10/500530
DT15 Rec'd PCT/PTO 01 JUL 2004
WO 03/055905 PCT/EP02/14902

Novel Compounds

FIELD OF THE INVENTION

This invention relates to polynucleotides, (herein referred to as "BASB231 polynucleotide(s)"), polypeptides encoded by them (referred to herein as "BASB231" or "BASB231 polypeptide(s)"), recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including vaccines against bacterial infections. In a further aspect, the invention relates to diagnostic assays for detecting infection of certain pathogens.

BACKGROUND OF THE INVENTION

Haemophilus influenzae is a non-motile Gram negative bacterium. Man is its only natural host.

H. influenzae isolates are usually classified according to their polysaccharide capsule. Six different capsular types designated a through f have been identified. Isolates that fail to agglutinate with antisera raised against one of these six serotypes are classified as non typeable, and do not express a capsule.

The *H. influenzae* type b is clearly different from the other types in that it is a major cause of bacterial meningitis and systemic diseases. non typeable *H. influenzae* (NTHi) are only occasionally isolated from the blood of patients with systemic disease.

NTHi is a common cause of pneumonia, exacerbation of chronic bronchitis, sinusitis and otitis media.

Otitis media is an important childhood disease both by the number of cases and its potential sequelae. More than 3.5 millions cases are recorded every year in the United States, and it is estimated that 80 % of children have experienced at least one episode of otitis before reaching the age of 3 (1). Left untreated, or becoming chronic, this disease may lead to hearing loss that can be temporary (in the case of fluid accumulation in the

middle ear) or permanent (if the auditive nerve is damaged). In infants, such hearing losses may be responsible for delayed speech learning.

Three bacterial species are primarily isolated from the middle ear of children with otitis media: *Streptococcus pneumoniae*, NTHi and *M. catarrhalis*. These are present in 60 to 90 % of cases. A review of recent studies shows that *S. pneumoniae* and NTHi each represent about 30 %, and *M. catarrhalis* about 15 % of otitis media cases (2). Other bacteria can be isolated from the middle ear (*H. influenzae* type B, *S. pyogenes*, ...) but at a much lower frequency (2 % of the cases or less).

Epidemiological data indicate that, for the pathogens found in the middle ear, the colonization of the upper respiratory tract is an absolute prerequisite for the development of an otitis; other factors are however also required to lead to the disease (3-9). These are important to trigger the migration of the bacteria into the middle ear via the Eustachian tubes, followed by the initiation of an inflammatory process. These other factors are unknown todate. It has been postulated that a transient anomaly of the immune system following a viral infection, for example, could cause an inability to control the colonization of the respiratory tract (5). An alternative explanation is that the exposure to environmental factors allows a more important colonization of some children, who subsequently become susceptible to the development of otitis media because of the sustained presence of middle ear pathogens (2).

Various proteins of *H. influenzae* have been shown to be involved in pathogenesis or have been shown to confer protection upon vaccination in animal models.

Adherence of NTHi to human nasopharygeal epithelial cells has been reported (10). Apart from fimbriae and pili (11-15), many adhesins have been identified in NTHi. Among them, two surface exposed high-molecular-weight proteins designated HMW1 and HMW2 have been shown to mediate adhesion of NTHi to epithelial cells (16). Another family of high molecular weight proteins has been identified in NTHi strains that lack proteins belonging to HMW1/HMW2 family. The NTHi 115 kDa Hia protein

(17) is highly similar to the Hsf adhesin expressed by *H. influenzae* type b strains (18). Another protein, the Hap protein shows similarity to IgA1 serine proteases and has been shown to be involved in both adhesion and cell entry (19).

Five major outer membrane proteins (OMP) have been identified and numerically numbered.

Original studies using *H.influenzae* type b strains showed that antibodies specific for P1 and P2 protected infant rats from subsequent challenge (20-21). P2 was found to be able to induce bactericidal and opsonic antibodies, which are directed against the variable regions present within surface exposed loop structures of this integral OMP (22-23). The lipoprotein P4 also could induce bactericidal antibodies (24).

P6 is a conserved peptidoglycan-associated lipoprotein making up 1-5 % of the outer membrane (25). Later a lipoprotein of about the same mol. wt. was recognized, called PCP (P6 crossreactive protein) (26). A mixture of the conserved lipoproteins P4, P6 and PCP did not reveal protection as measured in a chinchilla otitis-media model (27). P6 alone appears to induce protection in the chinchilla model (28).

P5 has sequence homology to the integral *Escherichia coli* OmpA (29-30). P5 appears to undergo antigenic drift during persistent infections with NTHi (31). However, conserved regions of this protein induced protection in the chinchilla model of otitis media.

In line with the observations made with gonococci and meningococci, NTHi expresses a dual human transferrin receptor composed of TbpA and TbpB when grown under iron limitation. Anti-TbpB protected infant rats. (32). Hemoglobin / haptoglobin receptors have also been described for NTHi (33). A receptor for Haem: Hemopexin has also been identified (34). A lactoferrin receptor is also present in NTHi, but is not yet characterized (35).

A 80kDa OMP, the D15 surface antigen, provides protection against NTHi in a mouse challenge model. (36). A 42kDa outer membrane lipoprotein,LPD is conserved amongst *Haemophilus influenzae* and induces bactericidal antibodies (37). A minor 98kDa OMP (38), was found to be a protective antigen, this OMP may very well be one of the Fe-limitation inducible OMPs or high molecular weight adhesins that have been characterized. *H. influenzae* produces IgA1-protease activity (39). IgA1-proteases of NTHi reveals a high degree of antigenic variability (40).

Another OMP of NTHi, OMP26, a 26-kDa protein has been shown to enhance pulmonary clearance in a rat model (41). The NTHi HtrA protein has also been shown to be a protective antigen. Indeed, this protein protected Chinchilla against otitis media and protected infant rats against *H. influenzae* type b bacteremia (42)

Background References

1. Klein, JO (1994) Clin.Inf.Dis 19:823
2. Murphy, TF (1996) Microbiol.Rev. 60:267
3. Dickinson, DP et al. (1988) J. Infect.Dis. 158:205
4. Faden, HL et al. (1991) Ann.Otorhinol.Laryngol. 100:612
5. Faden, HL et al (1994) J. Infect.Dis. 169:1312
6. Leach, AJ et al. (1994) Pediatr.Infect.Dis.J. 13:983
7. Prellner, KP et al. (1984) Acta Otolaryngol. 98:343
8. Stenfors, L-E and Raisanen, S. (1992) J.Infect.Dis. 165:1148
9. Stenfors, L-E and Raisanen, S. (1994) Acta Otolaryngol. 113:191
10. Read, RC. et al. (1991) J. Infect. Dis. 163:549
11. Brinton, CC. et al. (1989) Pediatr. Infect. Dis. J. 8:S54
12. Kar, S. et al. (1990) Infect. Immun. 58:903
13. Gildorf, JR. et al. (1992) Infect. Immun. 60:374
14. St. Geme, JW et al. (1991) Infect. Immun. 59:3366
15. St. Geme, JW et al. (1993) Infect. Immun. 61: 2233
16. St. Geme, JW. et al. (1993) Proc. Natl. Acad. Sci. USA 90:2875
17. Barenkamp, SJ. et JW St Geme (1996) Mol. Microbiol. (In press)

18. St. Geme, JW. et al. (1996) J. Bact. 178:6281
19. St. Geme, JW. et al. (1994) Mol. Microbiol. 14:217
20. Loeb, MR. et al. (1987) Infect. Immun. 55:2612
21. Musson, RS. Jr. et al. (1983) J. Clin. Invest. 72:677
22. Haase, EM. et al. (1994) Infect. Immun. 62:3712
23. Troelstra, A. et al. (1994) Infect. Immun. 62:779
24. Green, BA. et al. (1991) Infect.Immun.59:3191
25. Nelson, MB. et al. (1991) Infect. Immun. 59:2658
26. Deich, RM. et al. (1990) Infect. Immun. 58:3388
27. Green, BA. et al. (1993) Infect.immun. 61:1950
28. Demaria, TF. et al. (1996) Infect. Immun. 64:5187
29. Miyamoto, N., Bakaletz, LO (1996) Microb. Pathog. 21:343
30. Munson, RS.j.r. et al. (1993) Infect. Immun. 61:1017
31. Duim, B. et al. (1997) Infect. Immun. 65:1351
32. Loosmore, SM. et al(1996) Mol.Microbiol. 19:575
33. Maciver, I. et al. (1996) Infect. Immun. 64:3703
34. Cope, LD. et al. (1994) Mol.Microbiol. 13:868
35. Schryvers, AB. et al. (1989) J. Med. Microbiol. 29:121
36. Flack, FS. et al. (1995) Gene 156:97
37. Akkoyunlu, M. et al. (1996) Infect. Immun. 64:4586
38. Kimura, A. et al. (1985) Infect. Immun. 47:253
39. Mulks, MH. et Shoberg, RJ (1994) Meth. Enzymol. 235:543
40. Lomholt, H. Alphen, Lv, Kilian, M. (1993) Infect. Immun. 61:4575
41. Kyd, J.M. and Cripps, A.W. (1998) Infect. Immun. 66:2272
42. Loosmore, S.M. et al. (1998) Infect. Immun. 66:899

The frequency of NTHi infections has risen dramatically in the past few decades. This phenomenon has created an unmet medical need for new anti-microbial agents, vaccines, drug screening methods and diagnostic tests for this organism. The present invention aims to meet that need.

SUMMARY OF THE INVENTION

The present invention relates to BASB231, in particular BASB231 polypeptides and BASB231 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including prevention and treatment of microbial diseases, amongst others. In a further aspect, the invention relates to diagnostic assays for detecting diseases associated with microbial infections and conditions associated with such infections, such as assays for detecting expression or activity of BASB231 polynucleotides or polypeptides.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

DESCRIPTION OF THE INVENTION

The invention relates to BASB231 polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of BASB231 of non typeable *H. influenzae*.

The invention relates especially to BASB231 polynucleotides and encoded polypeptides listed in table 1. Those polynucleotides and encoded polypeptides have the nucleotide and amino acid sequences set out in SEQ ID NO:1 to SEQ ID NO:74 as described in table 1.

<u>Table 1</u>

Name	Length (nT)	Length (aa)	SEQ ID nucl.	SEQ ID prot.	Description
Orf1	453	150	1	2	LOS biosynthesis enzyme lbga, *Haemophilus ducreyi* (62%)
Orf2	1032	343	3	4	Putative d-glycero-d-manno-heptosyl transferase, *Actinobacillus pleuropneumoniae* (51%)
Orf3	813	270	5	6	Formamidopyrimidine-dna glycosylase, *Haemophilus*

					influenzae (74%)
Orf4	726	241	7	8	Molybdenum ABC transporter, periplasmic molybdate-binding protein, *Deinococcus radiodurans* (26%)
Orf5	741	246	9	10	ABC transporter, *Haemophilus influenzae* (38%)
Orf6	1023	340	11	12	ABC transporter, *Haemophilus influenzae* (45%)
Orf7	942	313	13	14	ABC transporter, *Haemophilus influenzae* (56%)
Orf8	558	185	15	16	Invasin precursor (YadA c-term), *Yersinia enterocolitica* (27%")
Orf9	2373	790	17	18	DNA methylase hsdm, *Vibrio cholerae* (70%)
Orf10	818	272	19	20	Leucyl tRNA synthetase, *Borrelia burgdorferi* (28%)
Orf11	636	211	21	22	ATP dependant DNA helicase, *Deinococcus radiodurans* (37%)
Orf12	1257	418	23	24	Type I restriction-modification system (s subunit), *Caulobacter crescentus* (29%)
Orf13	3027	1008	25	26	Type I restriction enzyme hsdr, *Vibrio cholerae* (65%)
Orf14	2052	683	27	28	Probable aaa family atpase, *Campylobacter jejuni* (33%)
Orf15	975	324	29	30	No homology with known protein
Orf16'	744	247	31	32	Hypothetical 29.0 kd protein, *Aquifex aeolicus* (24%)
Orf17	846	271	33	34	Hypothetical 27.0 kd protein, *Aquifex aeolicus* (30%)
Orf18	273	90	35	36	Cell division protein ftsk (C-term), *Escherichia coli* (46%)
Orf19	1023	340	37	38	Putative dna-binding protein, *Neisseria meningitidis* (45%)
Orf20	711	236	39	40	Hypothetical 22.9 kd protein, *Actinobacillus actinomycetemcomitans* (79%)
Orf21	456	151	41	42	Yors protein, *Bacillus subtilis* (26%)
Orf22	441	146	43	44	Phosphate transport atp-binding protein pstb homolog, *Mycoplasma genitalium* (24%)
Orf23	642	213	45	46	No homology with known protein
Orf24	1344	447	47	48	Type I restriction protein, *Haemophilus influenzae* (40%)
Orf25	1995	664	49	50	Hypothetical 84.7 kda protein, *Thermotoga maritima* (25%)
Orf26	1155	384	51	52	Anticodon nuclease, *Neisseria meningitidis* (61%)
Orf27	999	332	53	54	wkue. gp8 protein, *wolbachia* sp. (40 %)
Orf28	819	272	55	56	Putative transposase protein, *Rhizobium meliloti* (40%)
Orf29	333	110	57	58	Partial sequence of *Bacteriophage if1*. orf348 (35%)
Orf30	261	86	59	60	Putative cytoplasmic protein, *Salmonella typhimurium* lt2 (27%)
Orf31	927	308	61	62	Tryptophan 2-monooxygenase, *Agrobacterium*

					tumefaciens (29%)
Orf32	315	104	63	64	Modification methylase bepi, *Brevibacterium epidermidis* (51%)
Orf33	1464	487	65	66	PTS permease for n-acetylglucosamine and Glucose, *Vibrio furnissii* (71%)
Orf34	888	295	67	68	Putative lysr-family transcriptional regulator, *Neisseria meningitidis* (91%)
Orf35	843	280	69	70	Hypothetical 118.9 kda protein, *Plasmodium falciparum* (19%)
Orf36	393	130	71	72	tiorf34 protein, *Agrobacterium tumefaciens* (ti plasmid ptit37) (25%)
Orf37	675	224	73	74	Modification methylase bepi, *Brevibacterium epidermidis* (55%)

BASB231 polypeptides and polynucleotides are specific to non typeable *H. influenzae* (they are not present in *H. influenzae* Rd strain), and are thus particularly useful in the ntHi diagnostic field, as a whole host of ntHi-specific DNA probes and ntHi-specific enzyme functionalities may be used to detect the presence of ntHi in a sample as distinct from encapsuated Hi strains.

In addition, the availability of the above sequences allows: a) the upregulation or downregulation (i.e. knock-out of functional expression) of any of the above genes to create an ntHi strain with novel characteristics; b) the insertion and expression of any of the above genes in a non-ntHi host to introduce a ntHi-specific functionality into said host; and c) the purification of an ntHi-specific enzyme from the above list for performing in vitro reactions. To knock-out a gene, the gene (or a portion thereof) may be deleted, or may have an insertion or other mutation, or may have its promoter removed or replaced, such that expression of a gene product with the wild-type functionality is substantially (preferably completely) switched off. For instance Orf1 encodes a Lipo-oligosaccharide (LOS) biosynthesis enzyme (responsible for adding sugar groups to the antigenic ntHi-specific LOS molecule). With the Orf1 gene and protein sequences a skilled person will readily be able to ensure the above enzyme is either constitutively expressed or permanently switched off in a mutant ntHi strain in order to obtain a more consistent or a different LOS structure (respectively) which may be advantageously used for vaccine puroposes (either as LOS

complexed with ntHi outer membrane, or as purified LOS). In addition the enzyme may be isolated or recombinantly produced for its specific function to be used in vitro to produce novel synthetic oligosaccharide structures.

It is understood that sequences recited in the Sequence Listing below as "DNA" represent an exemplification of one embodiment of the invention, since those of ordinary skill will recognize that such sequences can be usefully employed in polynucleotides in general, including ribopolynucleotides.

The sequences of the BASB231 polynucleotides are set out in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73. SEQ Group 1 refers herein to any one of the polynucleotides set out in SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71 or 73.

The sequences of the BASB231 encoded polypeptides are set out in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72. SEQ Group 2 refers herein to any one of the encoded polypeptides set out in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70 or 72.

Polypeptides

In one aspect of the invention there are provided polypeptides of non typeable *H. influenzae* referred to herein as "BASB231" and "BASB231 polypeptides" as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

The present invention further provides for:

(a) an isolated polypeptide which comprises an amino acid sequence which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% or exact identity, to that of any sequence of SEQ Group 2;

(b) a polypeptide encoded by an isolated polynucleotide comprising a polynucleotide sequence which has at least 85% identity, preferably at least 90% identity, more preferably

at least 95% identity, even more preferably at least 97-99% or exact identity to any sequence of SEQ Group 1 over the entire length of the selected sequence of SEQ Group 1; or

(c) a polypeptide encoded by an isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% or exact identity, to the amino acid sequence of any sequence of SEQ Group 2.

The BASB231 polypeptides provided in SEQ Group 2 are the BASB231 polypeptides from non typeable *H. influenzae* strain ATCC PTA-1816.

The invention also provides an immunogenic (or enzymatically functional) fragment of a BASB231 polypeptide, that is, a contiguous portion of the BASB231 polypeptide which has the same or substantially the same immunogenic activity (or enzymatic activity) as the polypeptide comprising the corresponding amino acid sequence selected from SEQ Group 2 ; That is to say, the fragment (if necessary when coupled to a carrier) is capable of raising an immune response which recognises the BASB231 polypeptide (or can perform the same enzymatic function as the BASB231 polypeptide). Such an immunogenic (or enzymatically functional) fragment may include, for example, the BASB231 polypeptide lacking an N-terminal leader sequence, and/or a transmembrane domain and/or a C-terminal anchor domain. In a preferred aspect the immunogenic (or enzymatically functional) fragment of BASB231 according to the invention comprises substantially all of the extracellular domain of a polypeptide which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, most preferably at least 97-99% identity, to that a sequence selected from SEQ Group 2 over the entire length of said sequence.

A fragment is a polypeptide having an amino acid sequence that is entirely the same as part but not all of any amino acid sequence of any polypeptide of the invention. As with BASB231 polypeptides, fragments may be "free-standing," or comprised within a larger

polypeptide of which they form a part or region, most preferably as a single continuous region in a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of an amino acid sequence selected from SEQ Group 2 or of variants thereof, such as a continuous series of residues that includes an amino- and/or carboxyl-terminal amino acid sequence. Degradation forms of the polypeptides of the invention produced by or in a host cell, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Further preferred fragments include an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from an amino acid sequence selected from SEQ Group 2 or an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids truncated or deleted from an amino acid sequence selected from SEQ Group 2 .

Still further preferred fragments are those which comprise a B-cell or T-helper epitope, for example those fragments/peptides readily determined from the SEQ Group 2 sequences by well known prediction algorithms.

The B-cell epitopes of a protein are mainly localized at its surface. To predict B-cell epitopes of BASB231 polypeptides two methods can be combined: 2D-structure prediction and antigenic index prediction. The 2D-structure prediction can be made using the Chou Fasman method (from Chou PY and Fasman GD, Biochemistry, vol 13(2), pp 222-245, 1974)and the Gor method (from Garnier J, Osguthorpe DJ and Robson B, J Mol biol vol 120(1), pp97-120, 1978). The antigenic index can be calculated on the basis of the method described by Jameson and Wolf (CABIOS 4:181-

186 [1988]). The parameters used in this program are the antigenic index and the minimal length for an antigenic peptide. An antigenic index of 0.9 for a minimum of 5 consecutive amino acids is preferably used as threshold in the program. Peptides comprising potential B-cell epitopes can be useful (preferably conjugated or recombinantly joined to a larger protein) in a vaccine composition for the prevention of ntHi infections, and typically comprise 5 or more (e.g. 6, 7, 8, 9, 10, 11, 12, 15 or 20) contiguous amino acids from the BASB231 polypeptide sequence which can elicit an immune response in a host against the BASB231 polypeptide.

T-helper cell epitopes are peptides bound to HLA class II molecules and recognized by T-helper cells. The prediction of useful T-helper cell epitopes of BASB231 polypeptide is preferably based on the TEPITOPE method described by Sturniolo at al. (Nature Biotech. 17: 555-561 [1999]). Peptides comprising potential T-cell epitopes can be useful (preferably conjugated to peptides, polypeptides or polysaccharides) for vaccine purposes, and typically comprise 5 or more (e.g. 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 23, 26 or 30) contiguous amino acids from the BASB231 polypeptide sequence which preserve an effective T-helper epitope from BASB231 polypeptides.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these fragments may be employed as intermediates for producing the full-length polypeptides of the invention.

Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

The polypeptides, or immunogenic (or enzymatically functional) fragments, of the invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of

exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

In one aspect, the invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa.

Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

The proteins may be chemically conjugated, or expressed as recombinant fusion proteins allowing increased levels to be produced in an expression system as compared to non-fused protein. The fusion partner may assist in providing T helper epitopes (immunological fusion partner), preferably T helper epitopes recognised by humans, or assist in expressing the protein (expression enhancer) at higher yields than the native recombinant protein. Preferably the fusion partner will be both an immunological fusion partner and expression enhancing partner.

Fusion partners include protein D from *Haemophilus influenzae* and the non-structural protein from influenza virus, NS1 (hemagglutinin). Another fusion partner is the protein known as Omp26 (WO 97/01638). Another fusion partner is the protein known as LytA. Preferably the C terminal portion of the molecule is used. LytA is derived from *Streptococcus pneumoniae* which synthesize an N-acetyl-L-alanine amidase, amidase LytA, (coded by the *lytA* gene {Gene, 43 (1986) page 265-272}) an autolysin that

specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LytA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E.coli* C-LytA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LytA fragment at its amino terminus has been described {Biotechnology: 10, (1992) page 795-798}. It is possible to use the repeat portion of the LytA molecule found in the C terminal end starting at residue 178, for example residues 188 - 305.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

It is most preferred that a polypeptide of the invention is derived from non typeable *H. influenzae*, however, it may preferably be obtained from other organisms of the same taxonomic genus. A polypeptide of the invention may also be obtained, for example, from organisms of the same taxonomic family or order.

Polynucleotides

It is an object of the invention to provide polynucleotides that encode BASB231 polypeptides, particularly polynucleotides that encode the polypeptides herein designated BASB231.

In a particularly preferred embodiment of the invention the polynucleotides comprise a region encoding BASB231 polypeptides comprising sequences set out in SEQ Group 1 which include full length gene, or a variant thereof.

The BASB231 polynucleotides provided in SEQ Group 1 are the BASB231 polynucleotides from non typeable *H. influenzae* strain ATCC PTA-1816.

As a further aspect of the invention there are provided isolated nucleic acid molecules encoding and/or expressing BASB231 polypeptides and polynucleotides, particularly non typeable *H. influenzae* BASB231 polypeptides and polynucleotides, including, for example, unprocessed RNAs, ribozyme RNAs, mRNAs, cDNAs, genomic DNAs, B- and Z-DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful polynucleotides and polypeptides, and variants thereof, and compositions comprising the same.

Another aspect of the invention relates to isolated polynucleotides, including at least one full length gene, that encodes a BASB231 polypeptide having a deduced amino acid sequence of SEQ Group 2 and polynucleotides closely related thereto and variants thereof.

In another particularly preferred embodiment of the invention relates to BASB231 polypeptide from non typeable *H. influenzae* comprising or consisting of an amino acid sequence selected from SEQ Group 2 or a variant thereof.

Using the information provided herein, such as a polynucleotide sequences set out in SEQ Group 1 , a polynucleotide of the invention encoding BASB231 polypeptides may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using non typeable *H. influenzae* strain3224A cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a polynucleotide sequence given in SEQ Group 1, typically a library of clones of chromosomal DNA of

non typeable *H. influenzae* strain 3224A in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent hybridization conditions. By sequencing the individual clones thus identified by hybridization with sequencing primers designed from the original polypeptide or polynucleotide sequence it is then possible to extend the polynucleotide sequence in both directions to determine a full length gene sequence. Conveniently, such sequencing is performed, for example, using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Direct genomic DNA sequencing may also be performed to obtain a full length gene sequence. Illustrative of the invention, each polynucleotide set out in SEQ Group 1 was discovered in a DNA library derived from non typeable *H. influenzae*.

Moreover, each DNA sequence set out in SEQ Group 1 contains an open reading frame encoding a protein having about the number of amino acid residues set forth in SEQ Group 2 with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known to those skilled in the art.

The polynucleotides of SEQ Group 1, between the start codon and the stop codon, encode respectively the polypeptides of SEQ Group 2. The nucleotide number of start codon and first nucleotide of stop codon are listed in table 2 for each polynucleotide of SEQ Group 1.

Table 2

Name	Start codon	1st nucleotide of Stop codon
Orf1	1	453
Orf2	1	1030
Orf3	1	811
Orf4	1	724

Orf5	1	739
Orf6	1	1021
Orf7	1	940
Orf8	1*	556
Orf9	1	2371
Orf10	1	816
Orf11	1	634
Orf12	1	1255
Orf13	1	3025
Orf14	1	2050
Orf15	1	973
Orf16	1*	742
Orf17	1	814
Orf18	1*	271
Orf19	1	1021
Orf20	1	709
Orf21	1	454
Orf22	1*	439
Orf23	1	642
Orf24	1	1342
Orf25	1	1993
Orf26	1*	1153
Orf27	1	997
Orf28	1	817
Orf29	1*	331
Orf30	1	259
Orf31	1	916
Orf32	1*	310
Orf33	1	1462
Orf34	1	886
Orf35	1*	841
Orf36	1*	391
Orf37	1	673

*It is not the start codon but it is the first nucleotide of the coding sequence

In a further aspect, the present invention provides for an isolated polynucleotide comprising or consisting of:

(a) a polynucleotide sequence which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% or

exact identity, to any polynucleotide sequence from SEQ Group 1 over the entire length of the polynucleotide sequence from SEQ Group 1; or

(b) a polynucleotide sequence encoding a polypeptide which has at least 85% identity, preferably at least 90% identity, more preferably at least 95% identity, even more preferably at least 97-99% or 100% exact identity, to any amino acid sequence selected from SEQ Group 2 , over the entire length of the amino acid sequence from SEQ Group 2.

A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than non typeable *H. influenzae*, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions (for example, using a temperature in the range of 45 – 65°C and an SDS concentration from 0.1 – 1%) with a labeled or detectable probe consisting of or comprising any sequence selected from SEQ Group 1 or a fragment thereof; and isolating a full-length gene and/or genomic clones containing said polynucleotide sequence.

The invention provides a polynucleotide sequence identical over its entire length to a coding sequence (open reading frame) set out in SEQ Group 1. Also provided by the invention is a coding sequence for a mature polypeptide or a fragment thereof, by itself as well as a coding sequence for a mature polypeptide or a fragment in reading frame with another coding sequence, such as a sequence encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence. The polynucleotide of the invention may also contain at least one non-coding sequence, including for example, but not limited to at least one non-coding 5' and 3' sequence, such as the transcribed but non-translated sequences, termination signals (such as rho-dependent and rho-independent termination signals), ribosome binding sites, Kozak sequences, sequences that stabilize mRNA, introns, and polyadenylation signals. The polynucleotide sequence may also comprise additional coding sequence encoding additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA 86:* 821-824 (1989), or an HA peptide

tag (Wilson *et al.*, *Cell 37:* 767 (1984), both of which may be useful in purifying polypeptide sequence fused to them. Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The nucleotide sequence encoding the BASB231 polypeptide of SEQ Group 2 may be identical to the corresponding polynucleotide encoding sequence of SEQ Group 1. The position of the first and last nucleotides of the encoding sequences of SEQ Goup 1 are listed in table 3. Alternatively it may be any sequence, which as a result of the redundancy (degeneracy) of the genetic code, also encodes a polypeptide of SEQ Group 2 .

Table 3

Name	Start codon	Last nucleotide encoding polypeptide
Orf1	1	452
Orf2	1	1029
Orf3	1	810
Orf4	1	723
Orf5	1	738
Orf6	1	1020
Orf7	1	939
Orf8	1*	555
Orf9	1	2370
Orf10	1	815
Orf11	1	633
Orf12	1	1254
Orf13	1	3024
Orf14	1	2049
Orf15	1	972
Orf16	1*	741
Orf17	1	813
Orf18	1*	270
Orf19	1	1020
Orf20	1	708
Orf21	1	453
Orf22	1*	438
Orf23	1	641
Orf24	1	1341
Orf25	1	1992
Orf26	1*	1152

Orf27	1	996
Orf28	1	816
Orf29	1*	330
Orf30	1	258
Orf31	1	915
Orf32	1*	309
Orf33	1	1461
Orf34	1	885
Orf35	1*	840
Orf36	1*	390
Orf37	1	672

*It is not the start codon but it is the first nucleotide of the coding sequence

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the non typeable *H. influenzae* BASB231 having an amino acid sequence set out in any of the sequences of SEQ Group 2 . The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, polynucleotides interrupted by integrated phage, an integrated insertion sequence, an integrated vector sequence, an integrated transposon sequence, or due to RNA editing or genomic DNA reorganization) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode variants of a polypeptide having a deduced amino acid sequence of any of the sequences of SEQ Group 2 . Fragments of polynucleotides of the invention may be used, for example, to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding BASB231 variants, that have the amino acid sequence of BASB231 polypeptide of any sequence from SEQ Group 2 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, modified, deleted and/or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of BASB231 polypeptide.

Further preferred embodiments of the invention are polynucleotides that are at least 85% identical over their entire length to a polynucleotide encoding BASB231 polypeptide having an amino acid sequence set out in any of the sequences of SEQ Group 2 , and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 90% identical over its entire length to a polynucleotide encoding BASB231 polypeptide and polynucleotides complementary thereto. In this regard, polynucleotides at least 95% identical over their entire length to the same are particularly preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments are polynucleotides encoding polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by a DNA sequence selected from SEQ Group 1.

In accordance with certain preferred embodiments of this invention there are provided polynucleotides that hybridize, particularly under stringent conditions, to BASB231 polynucleotide sequences, such as those polynucleotides of SEQ Group 1.

The invention further relates to polynucleotides that hybridize to the polynucleotide sequences provided herein. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the polynucleotides described herein. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization occurring only if there is at least 95% and preferably at least 97% identity between the sequences. A specific example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml of denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C.

Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein. Solution hybridization may also be used with the polynucleotide sequences provided by the invention.

Such polynucleotides preferably have at least 15 or 30 nucleotide residues or base pairs and may have at least 50 nucleotide residues or base pairs. Particularly preferred polynucleotides will have at least 20 nucleotide residues or base pairs and will have less than 30 nucleotide residues or base pairs. Most preferably these polynucleotides are contiguous polynucleotides from a BASB231 polynucleotide sequence. Such polynucleotides are particularly useful in diagnostic methods where the specific hybridisation of these polynucleotides to the ntHi genome can differentiate the presence of ntHi in a sample rather than that of encapsulated Hi strains.

The invention also provides a polynucleotide consisting of or comprising a polynucleotide sequence obtained by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in any of the sequences of SEQ Group 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in the corresponding sequence of SEQ Group 1 or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers fully described elsewhere herein.

As discussed elsewhere herein regarding polynucleotide assays of the invention, for instance, the polynucleotides of the invention, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding BASB231 and to isolate cDNA and genomic clones of other genes that have a high identity, particularly high sequence identity, to the BASB231 gene. Such probes generally will comprise at least 15 nucleotide residues or base pairs. Preferably, such probes will have at least 30 nucleotide residues or base pairs and may have at least 50 nucleotide residues or base pairs. Particularly preferred probes will have at least 20 nucleotide residues or base pairs and will have less than 30 nucleotide residues or base pairs.

A coding region of a BASB231 gene may be isolated by screening using a DNA sequence provided in SEQ Group 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

There are several methods available and well known to those skilled in the art to obtain full-length DNAs, or extend short DNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman, *et al.*, *PNAS USA 85:* 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the DNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using "nested" primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the selected gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length DNA constructed either by joining the product directly to the existing DNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for diseases, particularly human diseases, as further discussed herein relating to polynucleotide assays.

The polynucleotides of the invention that are oligonucleotides derived from a sequence of SEQ Group 1 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are

transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

For each and every polynucleotide of the invention there is provided a polynucleotide complementary to it. It is preferred that these complementary polynucleotides are fully complementary to each polynucleotide with which they are complementary.

A precursor protein, having a mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleotides, the term "N" may also be used in describing certain polynucleotides of the invention. "N" means that any of the four DNA or RNA nucleotides may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a nucleic acid that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature

protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

In accordance with an aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al., Hum Mol Genet* (1992) 1: 363, Manthorpe *et al., Hum. Gene Ther.* (1983) 4: 419), delivery of DNA complexed with specific protein carriers (Wu *et al., J Biol Chem.* (1989) 264: 16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS USA*, (1986) 83: 9551), encapsulation of DNA in various forms of liposomes (Kaneda *et al., Science* (1989) 243: 375), particle bombardment (Tang *et al., Nature* (1992) 356:152, Eisenbraun *et al., DNA Cell Biol* (1993) 12: 791) and *in vivo* infection using cloned retroviral vectors (Seeger *et al., PNAS USA* (1984) 81: 5849).

Vectors, Host Cells, Expression Systems

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

Recombinant polypeptides of the present invention may be prepared by processes well known in those skilled in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems that comprise a polynucleotide or polynucleotides of the present

invention, to host cells which are genetically engineered with such expression systems, and to the production of polypeptides of the invention by recombinant techniques.

For recombinant production of the polypeptides of the invention, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis, *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook, *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, conjugation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as cells of streptococci, staphylococci, enterococci, *E. coli*, streptomyces, cyanobacteria, *Bacillus subtilis*, *Neisseria meningitidis*, *Haemophilus influenzae* and *Moraxella catarrhalis*; fungal cells, such as cells of a yeast, *Kluveromyces*, *Saccharomyces*, *Pichia*, a basidiomycete, *Candida albicans* and *Aspergillus*; insect cells such as cells of *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293, CV-1 and Bowes melanoma cells; and plant cells, such as cells of a gymnosperm or angiosperm.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal-, episomal- and virus-derived vectors, for example, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, picornaviruses, retroviruses, and alphaviruses and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender

expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

In recombinant expression systems in eukaryotes, for secretion of a translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, ion metal affinity chromatography (IMAC) is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

The expression system may also be a recombinant live microorganism, such as a virus or bacterium. The gene of interest can be inserted into the genome of a live recombinant virus or bacterium. Inoculation and *in vivo* infection with this live vector will lead to *in vivo* expression of the antigen and induction of immune responses. Viruses and bacteria used for this purpose are for instance: poxviruses (e.g; vaccinia, fowlpox, canarypox), alphaviruses (Sindbis virus, Semliki Forest Virus, Venezuelian Equine Encephalitis Virus), adenoviruses, adeno-associated virus, picornaviruses (poliovirus, rhinovirus), herpesviruses (varicella zoster virus, etc), *Listeria*, *Salmonella* , *Shigella*, BCG, streptococci. These viruses and bacteria can be virulent, or attenuated in various ways in order to obtain live vaccines. Such live vaccines also form part of the invention.

Diagnostic, Prognostic, Serotyping and Mutation Assays

This invention is also related to the use of BASB231 polynucleotides and polypeptides of the invention for use as diagnostic reagents. Detection of BASB231 polynucleotides and/or polypeptides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of disease, staging of disease or response of an infectious organism to drugs. Eukaryotes, particularly mammals, and especially humans, particularly those infected or suspected to be infected with an organism comprising the BASB231 gene or protein, may be detected at the nucleic acid or amino acid level by a variety of well known techniques as well as by methods provided herein.

Polypeptides and polynucleotides for prognosis, diagnosis or other analysis may be obtained from a putatively infected and/or infected individual's bodily materials. Polynucleotides from any of these sources, particularly DNA or RNA, may be used directly for detection or may be amplified enzymatically by using PCR or any other amplification technique prior to analysis. RNA, particularly mRNA, cDNA and genomic DNA may also be used in the same ways. Using amplification, characterization of the species and strain of infectious or resident organism present in an individual, may be made by an analysis of the genotype of a selected polynucleotide of the organism. Deletions and insertions can be detected by a change in size of the amplified product in comparison to a genotype of a reference sequence selected from a related organism, preferably a different species of the same genus or a different strain of the same species. Point mutations can be identified by hybridizing amplified DNA to labeled BASB231 polynucleotide sequences. Perfectly or significantly matched sequences can be distinguished from imperfectly or more significantly mismatched duplexes by DNase or RNase digestion, for DNA or RNA respectively, or by detecting differences in melting temperatures or renaturation kinetics. Polynucleotide sequence differences may also be detected by alterations in the electrophoretic mobility of polynucleotide fragments in gels as compared to a reference sequence. This may be carried out with or without denaturing agents. Polynucleotide differences may also be detected by direct DNA or RNA sequencing. See, for example, Myers *et al.*, *Science, 230:* 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays,

such as RNase, V1 and S1 protection assay or a chemical cleavage method. See, for example, Cotton *et al., Proc. Natl. Acad. Sci., USA, 85:* 4397-4401 (1985).

In another embodiment, an array of oligonucleotides probes comprising BASB231 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of, for example, genetic mutations, serotype, taxonomic classification or identification. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see, for example, Chee *et al., Science, 274: 610* (1996)).

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:
(a) a polynucleotide of the present invention, preferably any of the nucleotide sequences of SEQ Group 1, or a fragment thereof ;
(b) a nucleotide sequence complementary to that of (a);
(c) a polypeptide of the present invention, preferably any of the polypeptides of SEQ Group 2 or a fragment thereof; or
(d) an antibody to a polypeptide of the present invention, preferably to any of the polypeptides of SEQ Group 2 .

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a Disease, among others.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of a polynucleotide of the invention, preferably any sequence of SEQ Group 1 , which is associated with a disease or pathogenicity will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, a prognosis of a course of disease, a determination of a stage of disease, or a susceptibility to a disease, which results from under-expression, over-expression or altered expression of the polynucleotide. Organisms, particularly infectious organisms, carrying

mutations in such polynucleotide may be detected at the polynucleotide level by a variety of techniques, such as those described elsewhere herein.

Cells from an organism carrying mutations or polymorphisms (allelic variations) in a polynucleotide and/or polypeptide of the invention may also be detected at the polynucleotide or polypeptide level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations in the RNA. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA, cDNA or genomic DNA may also be used for the same purpose, PCR. As an example, PCR primers complementary to a polynucleotide encoding BASB231 polypeptide can be used to identify and analyze mutations.

The invention further provides primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. These primers may be used for, among other things, amplifying BASB231 DNA and/or RNA isolated from a sample derived from an individual, such as a bodily material. The primers may be used to amplify a polynucleotide isolated from an infected individual, such that the polynucleotide may then be subject to various techniques for elucidation of the polynucleotide sequence. In this way, mutations in the polynucleotide sequence may be detected and used to diagnose and/or prognose the infection or its stage or course, or to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections caused by non typeable *H. influenzae*, comprising determining from a sample derived from an individual, such as a bodily material, an increased level of expression of polynucleotide having a sequence of any of the sequences of SEQ Group 1. Increased or decreased expression of BASB231 polynucleotide can be measured using any on of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting, spectrometry and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of BASB231 polypeptide compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of BASB231 polypeptide, in a sample derived from a host, such as a bodily material, are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis, antibody sandwich assays, antibody detection and ELISA assays.

The polynucleotides of the invention may be used as components of polynucleotide arrays, preferably high density arrays or grids. These high density arrays are particularly useful for diagnostic and prognostic purposes. For example, a set of spots each comprising a different gene, and further comprising a polynucleotide or polynucleotides of the invention, may be used for probing, such as using hybridization or nucleic acid amplification, using a probes obtained or derived from a bodily sample, to determine the presence of a particular polynucleotide sequence or related sequence in an individual. Such a presence may indicate the presence of a pathogen, particularly non-typeable *Haemophilus influenzae*, and may be useful in diagnosing and/or prognosing disease or a course of disease. A grid comprising a number of variants of any polynucleotide sequence of SEQ Group 1 is preferred. Also preferred is a number of variants of a polynucleotide sequence encoding any polypeptide sequence of SEQ Group 2 .

<u>Antibodies</u>

The polypeptides and polynucleotides of the invention or variants thereof, or cells expressing the same can be used as immunogens to produce antibodies immunospecific for such polypeptides or polynucleotides respectively. Alternatively, mimotopes, particularly peptide mimotopes, of epitopes within the polypeptide sequence may also be used as immunogens to produce antibodies immunospecific for the polypeptide of the invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

In certain preferred embodiments of the invention there are provided antibodies against BASB231 polypeptides or polynucleotides.

Antibodies generated against the polypeptides or polynucleotides of the invention can be obtained by administering the polypeptides and/or polynucleotides of the invention, or epitope-bearing fragments of either or both, analogues of either or both, or cells expressing either or both, to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature 256:* 495-497 (1975); Kozbor *et al., Immunology Today 4:* 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides or polynucleotides of this invention. Also, transgenic mice, or other organisms or animals, such as other mammals, may be used to express humanized antibodies immunospecific to the polypeptides or polynucleotides of the invention.

Alternatively, phage display technology may be utilized to select antibody genes with binding activities towards a polypeptide of the invention either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-BASB231 or from naive libraries (McCafferty, *et al.*, (1990), Nature 348, 552-554; Marks, *et al.*, (1992) *Biotechnology 10*, 779-783). The affinity of these antibodies can also be improved by, for example, chain shuffling (Clackson *et al.*, (1991) *Nature 352:* 628).

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides or polynucleotides of the invention to purify the polypeptides or polynucleotides by, for example, affinity chromatography.

Thus, among others, antibodies against BASB231 polypeptide or BASB231 polynucleotide may be employed to treat infections, particularly bacterial infections.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants form a particular aspect of this invention.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized," where the complimentarity determining region or regions of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones *et al.* (1986), *Nature* 321, 522-525 or Tempest *et al.*, (1991) *Biotechnology* 9, 266-273.

<u>Antagonists and Agonists - Assays and Molecules</u>

Polypeptides and polynucleotides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology 1(2):* Chapter 5 (1991).

The screening methods may simply measure the binding of a candidate compound to the polypeptide or polynucleotide, or to cells or membranes bearing the polypeptide or polynucleotide, or a fusion protein of the polypeptide by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide or polynucleotide, using detection systems appropriate to the cells comprising the polypeptide or polynucleotide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polypeptide

and/or constitutively expressed polypeptides and polynucleotides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide or polynucleotide, as the case may be. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide or polynucleotide of the present invention, to form a mixture, measuring BASB231 polypeptide and/or polynucleotide activity in the mixture, and comparing the BASB231 polypeptide and/or polynucleotide activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and BASB231 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists of the polypeptide of the present invention, as well as of phylogenetically and and/or functionally related polypeptides (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies that bind to and/or interact with a polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and/or polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of BASB231 polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bactericidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising BASB231 polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence

or the presence of a candidate molecule that may be a BASB231 agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the BASB231 polypeptide is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of BASB231 polypeptide are most likely to be good antagonists. Molecules that bind well and, as the case may be, increase the rate of product production from substrate, increase signal transduction, or increase chemical channel activity are agonists. Detection of the rate or level of, as the case may be, production of product from substrate, signal transduction, or chemical channel activity may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric, labeled substrate converted into product, a reporter gene that is responsive to changes in BASB231 polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for BASB231 agonists is a competitive assay that combines BASB231 and a potential agonist with BASB231 binding molecules, recombinant BASB231 binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. BASB231 can be labeled, such as by radioactivity or a colorimetric compound, such that the number of BASB231 molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include, among others, small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide and/or polypeptide of the invention and thereby inhibit or extinguish its activity or expression. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing BASB231 induced activities, thereby preventing the action or expression of BASB231 polypeptides and/or polynucleotides by excluding BASB231 polypeptides and/or polynucleotides from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem. 56:* 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of BASB231.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Each of the polynucleotide sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the polynucleotide sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide, agonist or antagonist of the invention to interfere with the initial physical interaction between a

pathogen or pathogens and a eukaryotic, preferably mammalian, host responsible for sequelae of infection. In particular, the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive and/or gram negative bacteria, to eukaryotic, preferably mammalian, extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block bacterial adhesion between eukaryotic, preferably mammalian, extracellular matrix proteins and bacterial BASB231 proteins that mediate tissue damage and/or; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

In accordance with yet another aspect of the invention, there are provided BASB231 agonists and antagonists, preferably bacteristatic or bactericidal agonists and antagonists.

The antagonists and agonists of the invention may be employed, for instance, to prevent, inhibit and/or treat diseases.

In a further aspect, the present invention relates to mimotopes of the polypeptide of the invention. A mimotope is a peptide sequence, sufficiently similar to the native peptide (sequentially or structurally), which is capable of being recognised by antibodies which recognise the native peptide; or is capable of raising antibodies which recognise the native peptide when coupled to a suitable carrier.

Peptide mimotopes may be designed for a particular purpose by addition, deletion or substitution of elected amino acids. Thus, the peptides may be modified for the purposes of ease of conjugation to a protein carrier. For example, it may be desirable for some chemical conjugation methods to include a terminal cysteine. In addition it may be desirable for peptides conjugated to a protein carrier to include a hydrophobic terminus distal from the conjugated terminus of the peptide, such that the free unconjugated end of the peptide remains associated with the surface of the carrier protein. Thereby presenting the peptide in a conformation which most closely resembles that of the peptide as found in the context of the whole native molecule. For example, the peptides

may be altered to have an N-terminal cysteine and a C-terminal hydrophobic amidated tail. Alternatively, the addition or substitution of a D-stereoisomer form of one or more of the amino acids may be performed to create a beneficial derivative, for example to enhance stability of the peptide.

Alternatively, peptide mimotopes may be identified using antibodies which are capable themselves of binding to the polypeptides of the present invention using techniques such as phage display technology (EP 0 552 267 B1). This technique, generates a large number of peptide sequences which mimic the structure of the native peptides and are, therefore, capable of binding to anti-native peptide antibodies, but may not necessarily themselves share significant sequence homology to the native polypeptide.

<u>Vaccines</u>

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal, preferably humans, which comprises inoculating the individual with BASB231 polynucleotide and/or polypeptide, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly non typeable *H. influenzae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector, sequence or ribozyme to direct expression of BASB231 polynucleotide and/or polypeptide, or a fragment or a variant thereof, for expressing BASB231 polynucleotide and/or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual, preferably a human, from disease, whether that disease is already established within the individual or not. One example of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a

ribozyme, a modified nucleic acid, a DNA/RNA hybrid, a DNA-protein complex or an RNA-protein complex.

A further aspect of the invention relates to an immunological composition that when introduced into an individual, preferably a human, capable of having induced within it an immunological response, induces an immunological response in such individual to a BASB231 polynucleotide and/or polypeptide encoded therefrom, wherein the composition comprises a recombinant BASB231 polynucleotide and/or polypeptide encoded therefrom and/or comprises DNA and/or RNA which encodes and expresses an antigen of said BASB231 polynucleotide, polypeptide encoded therefrom, or other polypeptide of the invention. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity and/or cellular immunity, such as cellular immunity arising from CTL or CD4+ T cells.

BASB231 polypeptide or a fragment thereof may be fused with co-protein or chemical moiety which may or may not by itself produce antibodies, but which is capable of stabilizing the first protein and producing a fused or modified protein which will have antigenic and/or immunogenic properties, and preferably protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Haemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, or any other relatively large co-protein which solubilizes the protein and facilitates production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system of the organism receiving the protein. The co-protein may be attached to either the amino- or carboxy-terminus of the first protein.

In a vaccine composition according to the invention, a BASB231 polypeptide and/or polynucleotide, or a fragment, or a mimotope, or a variant thereof may be present in a vector, such as the live recombinant vectors described above for example live bacterial vectors.

Also suitable are non-live vectors for the BASB231 polypeptide, for example bacterial outer-membrane vesicles or "blebs". OM blebs are derived from the outer membrane of the two-layer membrane of Gram-negative bacteria and have been documented in many Gram-negative bacteria (Zhou, L *et al.* 1998. *FEMS Microbiol. Lett.* 163:223-228) including *C. trachomatis* and *C. psittaci.* A non-exhaustive list of bacterial pathogens reported to produce blebs also includes: *Bordetella pertussis, Borrelia burgdorferi, Brucella melitensis, Brucella ovis, Esherichia coli, Haemophilus influenzae, Legionella pneumophila, Moraxella catarrhalis, Neisseria gonorrhoeae, Neisseria meningitidis, Pseudomonas aeruginosa and Yersinia enterocolitica.*

Blebs have the advantage of providing outer-membrane proteins in their native conformation and are thus particularly useful for vaccines. Blebs can also be improved for vaccine use by engineering the bacterium so as to modify the expression of one or more molecules at the outer membrane. Thus for example the expression of a desired immunogenic protein at the outer membrane, such as the BASB231 polypeptide, can be introduced or upregulated (e.g. by altering the promoter). Instead or in addition, the expression of outer-membrane molecules which are either not relevant (e.g. unprotective antigens or immunodominant but variable proteins) or detrimental (e.g. toxic molecules such as LPS, or potential inducers of an autoimmune response) can be downregulated. These approaches are discussed in more detail below.

The non-coding flanking regions of the BASB231 gene contain regulatory elements important in the expression of the gene. This regulation takes place both at the transcriptional and translational level. The sequence of these regions, either upstream or downstream of the open reading frame of the gene, can be obtained by DNA sequencing. This sequence information allows the determination of potential regulatory motifs such as the different promoter elements, terminator sequences, inducible sequence elements, repressors, elements responsible for phase variation, the shine-dalgarno sequence, regions with potential secondary structure involved in regulation, as well as other types of regulatory motifs or sequences. This sequence is a further aspect of the invention.

Furthermore, SEQ ID NO: 75 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs1, 2, 3, 4, 5, 6, 7, 8 and their non-coding flanking regions.
The non-coding flanking regions are located between the ORFs of SED ID NO: 75. The localisation of the ORFs of SED ID NO: 75 are listed in table 4.

Table 4:

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf1	90	542	+
Orf2	545	1576	+
Orf3	2391	1579	-
Orf4	3165	2440	-
Orf5	3915	3175	-
Orf6	4934	3912	-
Orf7	5881	4940	-
Orf6	6579*	6022	-

* It is not the start codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 76 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 9, 10, 11, 12, 13 and their non-coding flanking regions.
The non-coding flanking regions are located between the ORFs of SED ID NO: 76. The localisation of the ORFs of SED ID NO: 76 are listed in table 5.

Table 5

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf9	140	2512	+
Orf10	2695	3512	+
Orf11	3470	4104	+
Orf12	4270	5526	+
Orf13	5626	8652	+

Furthermore, SEQ ID NO: 77 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 14, 15, 16, 17, 18, 19, 20, 21, 22 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 77. The localisation of the ORFs of SED ID NO: 77 are listed in table 6.

Table 6

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf14	2110	54	-
Orf15	3161	2187	-
Orf16	3931*	3239	-
Orf17	4854	4039	-
Orf18	5123*	4851	-
Orf19	5246	6268	+
Orf20	7027	6317	-
Orf21	7467	7011	-
Orf22	7966*	7526	-

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 78 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 23, 24 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 78. The localisation of the ORFs of SED ID NO: 78 are listed in table 7.

Table 7

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf23	688	47	-
Orf24	2028	685	-

Furthermore, SEQ ID NO: 79 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 25 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 79. The localisation of the ORF of SED ID NO: 79 are listed in table 8.

Table 8

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand

Orf25	2205	211	-

Furthermore, SEQ ID NO: 80 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 26, 27 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 80. The localisation of the ORFs of SED ID NO: 80 are listed in table 9.

Table 9

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf26	34*	1182	+
Orf27	1187	2185	+

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 81 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 28, 29 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 81. The localisation of the ORFs of SED ID NO: 81 are listed in table 10.

Table 10

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf28	152	970	+
Orf29	1729*	1397	-

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 82 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORFs 30, 31, 32 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORFs of SED ID NO: 82. The localisation of the ORFs of SED ID NO: 82 are listed in table 11.

Table 11

Name	Position of the first nucleotide of	Position of the last nucleotide of stop	Strand

	start codon	codon	
Orf30	271	11	-
Orf31	1154	237	-
Orf32	1475*	1164	-

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 83 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 33 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 83. The localisation of the ORF of SED ID NO: 83 are listed in table 12.

Table 12

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf33	74	1537	+

Furthermore, SEQ ID NO: 84 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 34 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 84. The localisation of the ORF of SED ID NO: 84 are listed in table 13.

Table 13

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf34	82	969	+

Furthermore, SEQ ID NO: 85 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 35 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 83. The localisation of the ORF of SED ID NO: 85 are listed in table 13.

Table 13

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand

Orf35	1065*	223	-

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 86 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 36 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 86. The localisation of the ORF of SED ID NO: 86 are listed in table 14.

Table 14

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf36	254*	646	+

*It is not the first nucleotide of the strat codon, it is the first nucleotide of the coding sequence

Furthermore, SEQ ID NO: 87 contains the non typeable *Haemophilus influenzae* polynucleotide sequences not present in the HiRd genome and comprising the ORF 37 and their non-coding flanking regions.

The non-coding flanking regions are located between the ORF of SED ID NO: 87. The localisation of the ORF of SED ID NO: 87 are listed in table 15.

Table 15

Name	Position of the first nucleotide of start codon	Position of the last nucleotide of stop codon	Strand
Orf37	202*	876	+

This sequence information allows the modulation of the natural expression of the BASB231 gene. The upregulation of the gene expression may be accomplished by altering the promoter, the shine-dalgarno sequence, potential repressor or operator elements, or any other elements involved. Likewise, downregulation of expression can be achieved by similar types of modification. Alternatively, by changing phase variation sequences, the expression of the gene can be put under phase variation control, or it may be uncoupled from this regulation. In another approach, the expression of the gene can be put under the control of one or more inducible elements allowing regulated expression. Examples of such regulation include, but are not limited to, induction by temperature

shift, addition of inductor substrates like selected carbohydrates or their derivatives, trace elements, vitamins, co-factors, metal ions, etc.

Such modifications as described above can be introduced by several different means. The modification of sequences involved in gene expression can be carried out *in vivo* by random mutagenesis followed by selection for the desired phenotype. Another approach consists in isolating the region of interest and modifying it by random mutagenesis, or site-directed replacement, insertion or deletion mutagenesis. The modified region can then be reintroduced into the bacterial genome by homologous recombination, and the effect on gene expression can be assessed. In another approach, the sequence knowledge of the region of interest can be used to replace or delete all or part of the natural regulatory sequences. In this case, the regulatory region targeted is isolated and modified so as to contain the regulatory elements from another gene, a combination of regulatory elements from different genes, a synthetic regulatory region, or any other regulatory region, or to delete selected parts of the wild-type regulatory sequences. These modified sequences can then be reintroduced into the bacterium via homologous recombination into the genome. A non-exhaustive list of preferred promoters that could be used for up-regulation of gene expression includes the promoters porA, porB, lbpB, tbpB, p110, lst, hpuAB from *N. meningitidis* or *N. gonorroheae;* ompCD, copB, lbpB, ompE, UspA1; UspA2; TbpB from *M. Catarrhalis;* p1, p2, p4, p5, p6, lpD, tbpB, D15, Hia, Hmw1, Hmw2 from *H. influenzae.*

In one example, the expression of the gene can be modulated by exchanging its promoter with a stronger promoter (through isolating the upstream sequence of the gene, in vitro modification of this sequence, and reintroduction into the genome by homologous recombination). Upregulated expression can be obtained in both the bacterium as well as in the outer membrane vesicles shed (or made) from the bacterium.

In other examples, the described approaches can be used to generate recombinant bacterial strains with improved characteristics for vaccine applications. These can be, but are not limited to, attenuated strains, strains with increased expression of selected antigens,

strains with knock-outs (or decreased expression) of genes interfering with the immune response, strains with modulated expression of immunodominant proteins, strains with modulated shedding of outer-membrane vesicles.

Thus, also provided by the invention is a modified upstream region of the BASB231 gene, which modified upstream region contains a heterologous regulatory element which alters the expression level of the BASB231 protein located at the outer membrane. The upstream region according to this aspect of the invention includes the sequence upstream of the BASB231 gene. The upstream region starts immediately upstream of the BASB231 gene and continues usually to a position no more than about 1000 bp upstream of the gene from the ATG start codon. In the case of a gene located in a polycistronic sequence (operon) the upstream region can start immediately preceding the gene of interest, or preceding the first gene in the operon. Preferably, a modified upstream region according to this aspect of the invention contains a heterologous promotor at a position between 500 and 700 bp upstream of the ATG.

The use of the disclosed upstream regions to upregulate the expression of the BASB231 gene, a process for achieving this through homologous recombination (for instance as described in WO 01/09350 incorporated by reference herein), a vector comprising upstream sequence suitable for this purpose, and a host cell so altered are all further aspects of this invention.

Thus, the invention provides a BASB231 polypeptide, in a modified bacterial bleb. The invention further provides modified host cells capable of producing the non-live membrane-based bleb vectors. The invention further provides nucleic acid vectors comprising the BASB231 gene having a modified upstream region containing a heterologous regulatory element.

Further provided by the invention are processes to prepare the host cells and bacterial blebs according to the invention.

Also provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides and/or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof, which have been shown to encode non-variable regions of bacterial cell surface proteins, in polynucleotide constructs used in such genetic immunization experiments in animal models of infection with non typeable *H. influenzae*. Such experiments will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value, derived from the requisite organ of the animal successfully resisting or clearing infection, for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly non typeable *H. influenzae* infection, in mammals, particularly humans.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant polypeptide and/or polynucleotide of the invention together with a suitable carrier, such as a pharmaceutically acceptable carrier. Since the polypeptides and polynucleotides may be broken down in the stomach, each is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostatic compounds and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

The vaccine formulation of the invention may also include adjuvant systems for enhancing the immunogenicity of the formulation. Preferably the adjuvant system raises preferentially a TH1 type of response.

An immune response may be broadly distinguished into two extreme catagories, being a humoral or cell mediated immune responses (traditionally characterised by antibody and cellular effector mechanisms of protection respectively). These categories of response have been termed TH1-type responses (cell-mediated response), and TH2-type immune responses (humoral response).

Extreme TH1-type immune responses may be characterised by the generation of antigen specific, haplotype restricted cytotoxic T lymphocytes, and natural killer cell responses. In mice TH1-type responses are often characterised by the generation of antibodies of the IgG2a subtype, whilst in the human these correspond to IgG1 type antibodies. TH2-type immune responses are characterised by the generation of a broad range of immunoglobulin isotypes including in mice IgG1, IgA, and IgM.

It can be considered that the driving force behind the development of these two types of immune responses are cytokines. High levels of TH1-type cytokines tend to favour the induction of cell mediated immune responses to the given antigen, whilst high levels of TH2-type cytokines tend to favour the induction of humoral immune responses to the antigen.

The distinction of TH1 and TH2-type immune responses is not absolute. In reality an individual will support an immune response which is described as being predominantly TH1 or predominantly TH2. However, it is often convenient to consider the families of cytokines in terms of that described in murine CD4 +ve T cell clones by Mosmann and Coffman (*Mosmann, T.R. and Coffman, R.L. (1989) TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annual Review of Immunology, 7, p145-173*). Traditionally, TH1-type responses are associated with the production of the INF-γ and IL-2 cytokines by T-lymphocytes. Other cytokines often

directly associated with the induction of TH1-type immune responses are not produced by T-cells, such as IL-12. In contrast, TH2- type responses are associated with the secretion of IL-4, IL-5, IL-6 and IL-13.

It is known that certain vaccine adjuvants are particularly suited to the stimulation of either TH1 or TH2 - type cytokine responses. Traditionally the best indicators of the TH1:TH2 balance of the immune response after a vaccination or infection includes direct measurement of the production of TH1 or TH2 cytokines by T lymphocytes *in vitro* after restimulation with antigen, and/or the measurement of the IgG1:IgG2a ratio of antigen specific antibody responses.

Thus, a TH1-type adjuvant is one which preferentially stimulates isolated T-cell populations to produce high levels of TH1-type cytokines when re-stimulated with antigen *in vitro*, and promotes development of both CD8+ cytotoxic T lymphocytes and antigen specific immunoglobulin responses associated with TH1-type isotype.

Adjuvants which are capable of preferential stimulation of the TH1 cell response are described in International Patent Application No. WO 94/00153 and WO 95/17209.

3 De-O-acylated monophosphoryl lipid A (3D-MPL) is one such adjuvant. This is known from GB 2220211 (Ribi). Chemically it is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains and is manufactured by Ribi Immunochem, Montana. A preferred form of 3 De-O-acylated monophosphoryl lipid A is disclosed in European Patent 0 689 454 B1 (SmithKline Beecham Biologicals SA).

Preferably, the particles of 3D-MPL are small enough to be sterile filtered through a 0.22micron membrane (European Patent number 0 689 454).
3D-MPL will be present in the range of 10μg - 100μg preferably 25-50μg per dose wherein the antigen will typically be present in a range 2-50μg per dose.

Another preferred adjuvant comprises QS21, an Hplc purified non-toxic fraction derived from the bark of *Quillaja Saponaria Molina*. Optionally this may be admixed with 3 De-O-acylated monophosphoryl lipid A (3D-MPL), optionally together with an carrier.

The method of production of QS21 is disclosed in US patent No. 5,057,540.

Non-reactogenic adjuvant formulations containing QS21 have been described previously (WO 96/33739). Such formulations comprising QS21 and cholesterol have been shown to be successful TH1 stimulating adjuvants when formulated together with an antigen.

Further adjuvants which are preferential stimulators of TH1 cell response include immunomodulatory oligonucleotides, for example unmethylated CpG sequences as disclosed in WO 96/02555.

Combinations of different TH1 stimulating adjuvants, such as those mentioned hereinabove, are also contemplated as providing an adjuvant which is a preferential stimulator of TH1 cell response. For example, QS21 can be formulated together with 3D-MPL. The ratio of QS21 : 3D-MPL will typically be in the order of 1 : 10 to 10 : 1; preferably 1:5 to 5 : 1 and often substantially 1 : 1. The preferred range for optimal synergy is 2.5 : 1 to 1 : 1 3D-MPL: QS21.

Preferably a carrier is also present in the vaccine composition according to the invention. The carrier may be an oil in water emulsion, or an aluminium salt, such as aluminium phosphate or aluminium hydroxide.

A preferred oil-in-water emulsion comprises a metabolisible oil, such as squalene, alpha tocopherol and Tween 80. In a particularly preferred aspect the antigens in the vaccine composition according to the invention are combined with QS21 and 3D-MPL in such an emulsion. Additionally the oil in water emulsion may contain span 85 and/or lecithin and/or tricaprylin.

Typically for human administration QS21 and 3D-MPL will be present in a vaccine in the range of 1μg - 200μg, such as 10-100μg, preferably 10μg - 50μg per dose. Typically the oil in water will comprise from 2 to 10% squalene, from 2 to 10% alpha tocopherol and from 0.3 to 3% tween 80. Preferably the ratio of squalene: alpha tocopherol is equal to or less than 1 as this provides a more stable emulsion. Span 85 may also be present at a level of 1%. In some cases it may be advantageous that the vaccines of the present invention will further contain a stabiliser.

Non-toxic oil in water emulsions preferably contain a non-toxic oil, e.g. squalane or squalene, an emulsifier, e.g. Tween 80, in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline.

A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil in water emulsion is described in WO 95/17210.

While the invention has been described with reference to certain BASB231 polypeptides and polynucleotides, it is to be understood that this covers fragments of the naturally occurring polypeptides and polynucleotides, and similar polypeptides and polynucleotides with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant polypeptides or polynucleotides.

The present invention also provides a polyvalent vaccine composition comprising a vaccine formulation of the invention in combination with other antigens, in particular antigens useful for treating *otitis media*. Such a polyvalent vaccine composition may include a TH-1 inducing adjuvant as hereinbefore described.

In a preferred embodiment, the polypeptides, fragments and immunogens of the invention are formulated with one or more of the following groups of antigens: a) one or more pneumococcal capsular polysaccharides (either plain or conjugated to a carrier protein); b) one or more antigens that can protect a host against *M. catarrhalis* infection; c) one or

more protein antigens that can protect a host against *Streptococcus pneumoniae* infection; d) one or more further non typeable *Haemophilus influenzae* protein antigens; e) one or more antigens that can protect a host against RSV; and f) one or more antigens that can protect a host against influenza virus. Combinations with: groups a) and b); b) and c); b), d), and a) and/or c); b), d), e), f), and a) and/or c) are preferred. Such vaccines may be advantageously used as global otitis media vaccines.

The pneumococcal capsular polysaccharide antigens are preferably selected from serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F and 33F (most preferably from serotypes 1, 3, 4, 5, 6B, 7F, 9V, 14, 18C, 19F and 23F).

Preferred pneumococcal protein antigens are those pneumococcal proteins which are exposed on the outer surface of the pneumococcus (capable of being recognised by a host's immune system during at least part of the life cycle of the pneumococcus), or are proteins which are secreted or released by the pneumococcus. Most preferably, the protein is a toxin, adhesin, 2-component signal tranducer, or lipoprotein of *Streptococcus pneumoniae*, or fragments thereof. Particularly preferred proteins include, but are not limited to: pneumolysin (preferably detoxified by chemical treatment or mutation) [Mitchell *et al.* Nucleic Acids Res. 1990 Jul 11; 18(13): 4010 "Comparison of pneumolysin genes and proteins from *Streptococcus pneumoniae* types 1 and 2.", Mitchell *et al.* Biochim Biophys Acta 1989 Jan 23; 1007(1): 67-72 "Expression of the pneumolysin gene in *Escherichia coli*: rapid purification and biological properties.", WO 96/05859 (A. Cyanamid), WO 90/06951 (Paton et al), WO 99/03884 (NAVA)]; PspA and transmembrane deletion variants thereof (US 5804193 - Briles *et al.*); PspC and transmembrane deletion variants thereof (WO 97/09994 - Briles et al); PsaA and transmembrane deletion variants thereof (Berry & Paton, Infect Immun 1996 Dec;64(12):5255-62 "Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*"); pneumococcal choline binding proteins and transmembrane deletion variants thereof; CbpA and transmembrane deletion variants thereof (WO 97/41151; WO 99/51266); Glyceraldehyde–3-phosphate

– dehydrogenase (Infect. Immun. 1996 64:3544); HSP70 (WO 96/40928); PcpA (Sanchez-Beato et al. *FEMS Microbiol Lett* 1998, 164:207-14); M like protein, SB patent application No. EP 0837130; and adhesin 18627, SB Patent application No. EP 0834568. Further preferred pneumococcal protein antigens are those disclosed in WO 98/18931, particularly those selected in WO 98/18930 and PCT/US99/30390.

Preferred further non-typeable *H. influenzae* protein antigens include Fimbrin protein (US 5766608) and fusions comprising peptides therefrom (eg LB1 Fusion) (US 5843464 - Ohio State Research Foundation), OMP26, P6, protein D, TbpA, TbpB, Hia, Hmw1, Hmw2, Hap, and D15.

Preferred influenza virus antigens include whole, live or inactivated virus, split influenza virus, grown in eggs or MDCK cells, or Vero cells or whole flu virosomes (as described by R. Gluck, Vaccine, 1992, 10, 915-920) or purified or recombinant proteins thereof, such as HA, NP, NA, or M proteins, or combinations thereof.

Preferred RSV (Respiratory Syncytial Virus) antigens include the F glycoprotein, the G glycoprotein, the HN protein, or derivatives thereof.

Compositions, kits and administration

In a further aspect of the invention there are provided compositions comprising a BASB231 polynucleotide and/or a BASB231 polypeptide for administration to a cell or to a multicellular organism.

The invention also relates to compositions comprising a polynucleotide and/or a polypeptides discussed herein or their agonists or antagonists. The polypeptides and polynucleotides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to an individual. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide and/or

polynucleotide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides, polynucleotides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide and/or polynucleotide, such as the soluble form of a polypeptide and/or polynucleotide of the present invention, agonist or antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides, polynucleotides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or

intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, solutions, powders and the like.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted

using standard empirical routines for optimization, as is well understood in the art.

Sequence Databases, Sequences in a Tangible Medium, and Algorithms

Polynucleotide and polypeptide sequences form a valuable information resource with which to determine their 2- and 3-dimensional structures as well as to identify further sequences of similar homology. These approaches are most easily facilitated by storing the sequence in a computer readable medium and then using the stored data in a known macromolecular structure program or to search a sequence database using well known searching tools, such as the GCG program package.

Also provided by the invention are methods for the analysis of character sequences or strings, particularly genetic sequences or encoded protein sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, DNA, RNA and protein structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, codon usage analysis, nucleic acid base trimming, and sequencing chromatogram peak analysis.

A computer based method is provided for performing homology identification. This method comprises the steps of: providing a first polynucleotide sequence comprising the sequence of a polynucleotide of the invention in a computer readable medium; and comparing said first polynucleotide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

A computer based method is also provided for performing homology identification, said method comprising the steps of: providing a first polypeptide sequence comprising the sequence of a polypeptide of the invention in a computer readable medium; and comparing said first polypeptide sequence to at least one second polynucleotide or polypeptide sequence to identify homology.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

DEFINITIONS

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heine, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GAP program in the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1):* 387 (1984)), BLASTP, BLASTN (Altschul, S.F. et al., *J. Molec. Biol. 215:* 403-410 (1990), and FASTA(Pearson and Lipman Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988). The BLAST family of programs is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894;

Altschul, S., *et al.*, *J. Mol. Biol. 215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:
Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: BLOSSUM62 from Henikoff and Henikoff,
Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)
Gap Penalty: 8
Gap Length Penalty: 2
A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Parameters for polynucleotide comparison include the following:
Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)
Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50
Gap Length Penalty: 3
Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO:1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and

wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and $\bullet$ is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n. Alterations of polynucleotide sequences encoding the polypeptides of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequences of SEQ ID NO:1, that is it may be 100% identical, or it may include up to a certain integer number of nucleic acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of nucleic acids in SEQ ID NO:1

by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleic acids in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleic acid alterations, x_n is the total number of nucleic acids in SEQ ID NO:1, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., • is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n.

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to the polypeptide reference sequence of SEQ ID NO:2, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO:2 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and • is the symbol for

the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and $\bullet$ is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a.

"Individual(s)," when used herein with reference to an organism, means a multicellular eukaryote, including, but not limited to a metazoan, a mammal, an ovid, a bovid, a simian, a primate, and a human.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not “isolated,” but

the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA including single and double-stranded regions.

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Disease(s)" means any disease caused by or related to infection by a bacteria, including, for example, otitis media in infants and children, pneumonia in elderlies, sinusitis, nosocomial infections and invasive diseases, chronic otitis media with hearing loss, fluid

accumulation in the middle ear, auditive nerve damage, delayed speech learning, infection of the upper respiratory tract and inflammation of the middle ear.

EXAMPLES:

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1: Cloning of the BASB231 gene from non typeable *Haemophilus influenzae* strain 3224A.

Genomic DNA is extracted from the non typeable *Haemophilus influenzae* strain 3224A from 10^{10} bacterial cells using the QIAGEN genomic DNA extraction kit (Qiagen Gmbh). This material (1 µg) is then submitted to Polymerase Chain Reaction DNA amplification using two specific primers. A DNA fragment is obtained, digested by the suitable restriction endonucleases and inserted into the compatible sites of the pET cloning/expression vector (Novagen) using standard molecular biology techniques (Molecular Cloning, a Laboratory Manual, Second Edition, Eds: Sambrook, Fritsch & Maniatis, Cold Spring Harbor press 1989). Recombinant pET-BASB231 is then submitted to DNA sequencing using the Big Dyes kit (Applied biosystems) and analyzed on a ABI 373/A DNA sequencer in the conditions described by the supplier.

Example 2: Expression and purification of recombinant BASB231 protein in *Escherichia coli*.

The construction of the pET-BASB231 cloning/expression vector is described in Example 1. This vector harbours the BASB231 gene isolated from the non typeable *Haemophilus influenzae strain 3224A* in fusion with a stretch of 6 Histidine residues, placed under the control of the strong bacteriophage T7 gene 10 promoter. For expression study, this vector is introduced into the *Escherichia coli* strain Novablue (DE3) (Novagen), in which, the gene for the T7 polymerase is placed under the control of the isopropyl-beta-D thiogalactoside (IPTG)-regulatable *lac* promoter. Liquid cultures (100 ml) of the Novablue (DE3) [pET-BASB231] *E. coli* recombinant strain are grown at 37°C under

agitation until the optical density at 600nm (OD600) reached 0.6. At that time-point, IPTG is added at a final concentration of 1mM and the culture is grown for 4 additional hours. The culture is then centrifuged at 10,000 rpm and the pellet is frozen at -20°C for at least 10 hours. After thawing, the pellet is resuspended during 30 min at 25°C in buffer A (6M guanidine hydrochloride, 0.1M NaH2PO4, 0.01M Tris, pH 8.0), passed three-times through a needle and clarified by centrifugation (20000rpm, 15 min). The sample is then loaded at a flow-rate of 1ml/min on a Ni2+ -loaded Hitrap column (Pharmacia Biotech). After passsage of the flowthrough, the column is washed succesively with 40ml of buffer B (8M Urea, 0.1MNaH2PO4, 0.01M Tris, pH 8.0), 40ml of buffer C (8M Urea, 0.1MNaH2PO4, 0.01M Tris, pH 6.3). The recombinant protein BASB231/His6 is then eluted from the column with 30ml of buffer D (8M Urea, 0.1MNaH2PO4, 0.01M Tris, pH 6.3) containing 500mM of imidazole and 3ml-size fractions are collected. Highly enriched BASB231/His6 protein can be eluted from the column. This polypeptide is detected by a mouse monoclonal antibody raised against the 5-histidine motif. Moreover, the denatured, recombinant BASB231-His6 protein is solubilized in a solution devoid of urea. For this purpose, denatured BASB231-His6 contained in 8M urea is extensively dialyzed (2 hours) against buffer R (NaCl 150mM, 10mM NaH2PO4, Arginine 0.5M pH6.8) containing successively 6M, 4M, 2M and no urea. Alternatively, this polypeptide is purified under non-denaturing conditions using protocoles described in the Quiexpresssionist booklet (Qiagen Gmbh).

Example 3: Production of Antisera to Recombinant BASB231

Polyvalent antisera directed against the BASB231 protein are generated by vaccinating rabbits with the purified recombinant BASB231 protein. Polyvalent antisera directed against the BASB231 protein are also generated by vaccinating mice with the purified recombinant BASB231 protein. Animals are bled prior to the first immunization ("pre-bleed") and after the last immunization.

Anti-BASB231 protein titers are measured by an ELISA using purified recombinant BASB231 protein as the coating antigen. The titer is defined as mid-point titers

calculated by 4-parameter logistic model using the XL Fit software.The antisera are also used as the first antibody to identify the protein in a western blot as described in example 5 below.

Example 4: Immunological characterization: Surface exposure of BASB231

Anti-BASB231 protein titres are determined by an ELISA using formalin-killed whole cells of non typable *Haemophilus influenzae* (NTHi). The titer is defined as mid-point titers calculated by 4-parameter logistic model using the XL Fit software.

Example 5. Immunological Characterisation: Western Blot Analysis

Several strains of NTHi, as well as clinical isolates, are grown on Chocolate agar plates for 24 hours at 36°C and 5% CO_2. Several colonies are used to inoculate Brain Heart Infusion (BHI) broth supplemented by NAD and hemin, each at 10 µg/ml. Cultures are grown until the absorbance at 620nm is approximately 0.4 and cells are collected by centrifugation. Cells are then concentrated and solubilized in PAGE sample buffer. The solubilized cells are then resolved on 4-20% polyacrylamide gels and the separated proteins are electrophoretically transferred to PVDF membranes. The PVDF membranes are then pretreated with saturation buffer. All subsequent incubations are carried out using this pretreatment buffer.

PVDF membranes are incubated with preimmune serum or rabbit or mouse immune serum. PVDF membranes are then washed.

PVDF membranes are incubated with biotin-labeled sheep anti-rabbit or mouse Ig. PVDF membranes are then washed 3 times with wash buffer, and incubated with streptavidin-peroxydase. PVDF membranes are then washed 3 times with wash buffer and developed with 4-chloro-1-naphtol.

Example 6: Immunological characterization: Bactericidal Activity

Complement-mediated cytotoxic activity of anti-BASB231 antibodies is examined to determine the vaccine potential of BASB231 protein antiserum that is prepared as

described above. The activities of the pre-immune serum and the anti-BASB231 antiserum in mediating complement killing of NTHi are examined.

Strains of NTHi are grown on plates. Several colonies are added to liquid medium. Cultures are grown and collected until the A620 is approximately 0.4. After one wash step, the pellet is suspended and diluted.

Preimmune sera and the anti-BASB231 sera are deposited into the first well of a 96-wells plate and serial dilutions are deposited in the other wells of the same line. Live diluted NTHi is subsequently added and the mixture is incubated. Complement is added into each well at a working dilution defined beforehand in a toxicity assay.

Each test includes a complement control (wells without serum containing active or inactivated complement source), a positive control (wells containing serum with a know titer of bactericidal antibodies), a culture control (wells without serum and complement) and a serum control (wells without complement).

Bactericidal activity of rabbit or mice antiserum (50% killing of homologous strain) is measured.

Example 7: Presence of Antibody to BASB231 in Human Convalescent Sera

Western blot analysis of purified recombinant BASB231 is performed as described in Example 5 above, except that a pool of human sera from children infected by NTHi is used as the first antibody preparation.

Example 8: Efficacy of BASB231 vaccine: enhancement of lung clearance of NTHi in mice.

This mouse model is based on the analysis of the lung invasion by NTHi following a standard intranasal challenge to vaccinated mice.

Groups of mice are immunized with BASB231 vaccine. After the booster, the mice are challenged by instillation of bacterial suspension into the nostril under anaesthesia.

Mice are killed between 30 minutes and 24 hours after challenge and the lungs are removed aseptically and homogenized individually. The log10 weighted mean number of CFU/lung is determined by counting the colonies grown on agar plates after plating of dilutions of the homogenate. The arithmetic mean of the log10 weighted mean number of CFU/lung and the standard deviations are calculated for each group. Results are analysed statistically.

In this experiment groups of mice are immunized either with BASB231 or with a killed whole cells (kwc) preparation of NTHi or sham immunized.

Example 9: Inhibition of NTHi adhesion onto cells by anti-BASB231 antiserum.

This assay measures the capacity of anti BASB231 sera to inhibit the adhesion of NTHi bacteria to epithelial cells. This activity could prevent colonization of the nasopharynx by NTHi.

One volume of bacteria is incubated on ice with one volume of pre-immune or anti-BASB231 immune serum dilution. This mixture is subsequently added in the wells of a 24 well plate containing a confluent cells culture that is washed once with culture medium to remove traces of antibiotic. The plate is centrifuged and incubated.

Each well is then gently washed. After the last wash, sodium glycocholate is added to the wells. After incubation, the cell layer is scraped and homogenised. Dilutions of the homogenate are plated on agar plates and incubated. The number of colonies on each plate is counted and the number of bacteria present in each well calculated.

Deposited materials

A deposit of strain 3 (strain 3224A) has been deposited with the American Type Culture Collection (ATCC) on May 5 2000 and assigned deposit number PTA-1816.

The non typeable *Haemophilus influenzae* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited strain contains a full length BASB231 polynucleotide sequence.

The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of any polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit of the deposited strain has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The deposited strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strain is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112. A license may be required to make, use or sell the deposited strain, and compounds derived therefrom, and no such license is hereby granted.

Applicant's or agent's file reference number MJL/B45292	International application No.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 70 lines 1-22.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

AMERICAN TYPE CULTURE COLLECTION

Address of depositary institution *(including postal code and country)*

10801 UNIVERSITY BLVD, MANASSAS, VIRGINIA 20110-2209, UNITED STATES OF AMERICA

Date of deposit 5 May 2000	Accession Number PTA-1816

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

In respect of those designations where a European Patent is sought, a sample of the deposited microorganisms will be made available until the publication of the mention of the grant of the European Patent or until the date on which the application has been refused or withdrawn, only by issue of such a sample to an expert nominated by the person requesting the sample

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

SEQUENCE INFORMATION

BASB231 Polynucleotide and Polypeptide Sequences

SEQ ID NO:1 polynucleotide sequence of Orf1

GTGTGCTATGAGCCATTTATTTATTACCCAATGATGTGCAATGAAAAGATAGCGCGTGCTATTATTCTTG
AAGATGATGCGATTGTATCGCACGAATTCGAAGCAATTGTAAAAGACAGTTTGAAGAAAGTTTCAAAAAA
TGTTGAAATTTTATTTTATGATCATGGTAAAGCAAAAAGTTATTGCTGGAAAAAAACACTTGTCAAAAAT
TACCGTTTAGTTCACTATCGTAAACCCTCTAAAACGTCTAAACGTGCAATCATGTGTACAACAGCTTATT
TAATTACTTTATCTGGCGCTCAAAAACTCCTACAAATAGCCTATCCTATCCGTATGCCTGCTGACTACTT
AACTGGTGCTTTACAATTAACTGGACTAAAGGCTTATGGTGTTGAACCACCTTGTGTATTTAAAGGCGCA
ATTTCAGAAATTGATGCAATGGAGCAACGCTAA

SEQ ID NO:2 polypeptide sequence of Orf1

VCYEPFIYYPMMCNEKIARAIILEDDAIVSHEFEAIVKDSLKKVSKNVEILFYDHGKAKSYCWKKTLVKNYR
LVHYRKPSKTSKRAIMCTTAYLITLSGAQKLLQIAYPIRMPADYLTGALQLTGLKAYGVEPPCVFKGAISEI
DAMEQR.

SEQ ID NO:3 polynucleotide sequence of Orf2

ATGAAATTAAAAAATAAATTACAAATGTTAAGGTTGGGTCTAGGCAAATATTTCCTTGATAAAAAAAACG
GATTAAACAGAATAACAAATGTTCCTAGAAGCATCCTCTTCCTCCGCCAAGACGGAAAAATTGGGGATTA
TGTGGTGAGCTCATTTGTATTCCGTGAGATAAAAAAATTTAATCCCCACATTAAAATTGGTGTAATTTGT
ACCAAACAAAATGCTTATCTTTTTAAACAAAATCCATATATCGATCAACTTTACTATGTAAAAAAGAAAA
GTATTTTGGATTACATCAAATGTGGTCTAGCAATTCAAAAAGAACAATATGATTTAGTGATTGATCCGAC
GATTATGATTCGTAATCGCGATCTTTTACTTTTACGCTTAATCAATGCCAAGCATTATATTGGCTACCAA
AAAGCCAATTATGGTTTATTTAATATTAATCTGGAGGGACAATTTCACTTTTCGGAACTCTATAAACTCG
CCTTAGAAAAAGTGAATATTACGGTACAAGATATAAGCTATGACATCCCATTTGATAAGCAAAGTGCGGT
CGAAATTTCTGAATTTTTGCAGAAAAACCAACTAGAAAAGTATATTGCTATTAATTTTTATGGTGCTGCA
AGAATCAAAAAAGTAAACAATGACAACATCAAAAAATATTTAGATTATCTCACGCAAGTCCGCGGAGGAA
AAAAGCTGGTGCTATTAAGCTATCCTGAAGTAACAGAGAAATTAACACAATTGTCAGCCGATTATCCGCA
TATTTTTGTCCATCCAACAACCAAGATCTTTCATACCATTGAATTGATTCGCCACTGTGATCAATTAATC
TCTACAGACACGTCTACTGTACATATTGCTTCAGGTTTTAATAAACCAATTATTGGTATTTATAAAGAAG
ATCCTATTGCGTTTACACATTGGCAACCCAGAAGTCGGGCAGAAACGCACATACTTTTCTATAAAGAAAA
TATTAATGAGCTCTCACCTGAACAAATTGACCCTGCATGGCTTGTCAAATAG

SEQ ID NO:4 polypeptide sequence of Orf2

MKLKNKLQMLRLGLGKYFLDKKNGLNRITNVPRSILFLRQDGKIGDYVVSSFVFREIKKFNPHIKIGVICTK
QNAYLFKQNPYIDQLYYVKKKSILDYIKCGLAIQKEQYDLVIDPTIMIRNRDLLLLRLINAKHYIGYQKANY
GLFNINLEGQFHFSELYKLALEKVNITVQDISYDIPFDKQSAVEISEFLQKNQLEKYIAINFYGAARIKKVN
NDNIKKYLDYLTQVRGGKKLVLLSYPEVTEKLTQLSADYPHIFVHPTTKIFHTIELIRHCDQLISTDTSTVH
IASGFNKPIIGIYKEDPIAFTHWQPRSRAETHILFYKENINELSPEQIDPAWLVK.

SEQ ID NO:5 polynucleotide sequence of Orf3

ATGCCAGAATTACCTGAAGTTGAAACCACAAAAAATGGAATTAGCCCTTATCTTGAAGGGGCTATCATTG
AAAAAATTGTTGTTCGCCAACCGAAATTACGCTGGATGGTAAGCGAAGAATTAGCGCAAATTACACAACA

AAAAGTCATCGCATTAAGTCGCCGTGCGAAGTATTTAATTATCCAACTTGAAACAGGCTATATGATTGGA
CATTTAGGGATGTCAGGGTCATTGAGAGTTGTGGAGAAAGGGGATCTTATTGATAAACATGATCATCTTG
ATATCGTAGTGAATAACGGAAAAGTTGTGCGTTATAACGATCCTCGTCGTTTTGGAGCGTGGTTATGGAC
AGAGAAGTTGAACGAATTTCCTCTTTTTCTGAAATTAGGCCCAGAGCCTCTGTCTGAGGAATTTGATTCT
GATTACTTGTGGCAAAAAAGTCGTAAAAAACAGACCGCACTTAAAACTTTTTTAATGGATAATGCTGTCG
TCGTTGGCGTTGGGAATATCTATGCGAATGAAACGTTATTTCTTTGTAACCTACATCCGCAAAAAACAGC
AGGGAGTTTAACTAAGGCACAATGTGGGCAGTTAGTAGAACAAATAAAACAAGTGCTGTCTAACGCAATC
CAACAAGGTGGTACGACGCTAAAAGATTTTCTCCAACCGGATGGGCGTCCAGGCTATTTTGTCCAAGAAT
TGCGGGTTTATGGTAATAAGGATAAGCCTTGTCCAACATGTGGCACAAAAATAGAAAGTTTAGTGATAGG
GCAACGAAATAGTTTCTATTGCCCCAAGTGTCAGAAGAGATAA

SEQ ID NO:6 polypeptide sequence of Orf3

MPELPEVETTKNGISPYLEGAIIEKIVVRQPKLRWMVSEELAQITQQKVIALSRRAKYLIIQLETGYMIGHL
GMSGSLRVVEKGDLIDKHDHLDIVVNNGKVVRYNDPRRFGAWLWTEKLNEFPLFLKLGPEPLSEEFDSDYLW
QKSRKKQTALKTFLMDNAVVVGVGNIYANETLFLCNLHPQKTAGSLTKAQCGQLVEQIKQVLSNAIQQGGTT
LKDFLQPDGRPGYFVQELRVYGNKDKPCPTCGTKIESLVIGQRNSFYCPKCQKR.

SEQ ID NO:7 polynucleotide sequence of Orf4

ATGAGAATTTTAGCCGCAGGGAGTTTACGCCAGCCTTTTACGTTATGGCAACAAGCATTAATCCAACAGT
ATCACCTACAAGTCGAAATTGAATTTGGACCGGCGGGGTTGTTGTGCCAACGCATTGAGCAAGGGGAAAA
AGTGGATTTGTTTGCCTCTGCCAATGATGCGCATCTTAGGCATTTACAAGCGCGATATCCTCATATTCAA
CTTGTGCCTTTTGCTACAAATCGTTTATGTTTAATTGCAAAGAAATCGGTGATTACTCACCATGATGAGA
ATTGGTTGACATTATTGATGTCGCCCCACTTACGCTTAGGAGTATCGACACCTAAGGCAGATCCTTGTGG
AGATTATACTTTGGCATTATTTTCGAATATTGAAAAACGGCATATGGGCTATGGCTCGGAATTAAAAGAA
AAAGCAATGGCAATAGTTGGTGGTCCGGATTCTATCACTATTCCAACAGGACGAAATACCGCAGAGTGGC
TTTTTGAGCAGAATTATGCTGATCTTTTCATTGGTTATGCGAGTAATCATCAATCTTTGCGTCAGCATTC
TGATATTTGTGTTTTGGATATTCCTGATGAGTATAATGTGAGGGCGAACTATACATTAGCAGCTTTTACT
GCGGAAGCATTACGCCTTGTGGACTCCTTGCTTTGTTTGACTTGCGGACAAAAATATTTACGCGATTGCG
GCTTTTTGCCTGCCAATCATAGCTGA

SEQ ID NO:8 polypeptide sequence of Orf4

MRILAAGSLRQPFTLWQQALIQQYHLQVEIEFGPAGLLCQRIEQGEKVDLFASANDAHLRHLQARYPHIQLV
PFATNRLCLIAKKSVITHHDENWLTLLMSPHLRLGVSTPKADPCGDYTLALFSNIEKRHMGYGSELKEKAMA
IVGGPDSITIPTGRNTAEWLFEQNYADLFIGYASNHQSLRQHSDICVLDIPDEYNVRANYTLAAFTAEALRL
VDSLLCLTCGQKYLRDCGFLPANHS.

SEQ ID NO:9 polynucleotide sequence of Orf5

ATGAATGAATTGAGTTTAGATGCAGATAAGCTGTTATTTGGTTATGATAAGCCGTTGTATTTACCACTTACT
TTCCAATGTAAGAAAGGAGAGGTTATTTCGGTATTTGGAACAAATGGAAAAGGTAAAACCACATTATTGCAT
TCTCTTGCTCATGTGTTACCTGTTATGTCTGGACAGATTAGGCAACAAGGTCATATTGGTTTTGTGCCACAG
TCTTTTTCGTCGCCAGATTATCCCGTGTTAGAGATTGTTTTAATGGGGCGAGCAAGCAAAATTGGAGCATTT
AACTTACCAAGTAAAACGGATGAAACAGTCGCATTACAGATGTTGGCGTGCTTAGACATCCTGCATTTAGCT
GAGCGCAATATCAATATGCTTTCGGGCGGTCAACGCCAACTTGTGCTCATCGCTCGTGCACTTGCGACAGAA
TGTCAGGTCCTCATTTTAGATGAACCTACAGCAGCATTGGATGTTTATAATCAATAGCGTGTCTTACAACTT
ATACGTTTTCTTGCAACGGAACAAAAAATGACCATTATTTTTTCCACTCATGATCCTTATCACAGTTTATGT
GTGGCAGATAATGTGTTATTGCTATTGCCTAACCAACAATGGAAATATGGAATAGCCAGTCAAATTTTAACG
GAATCTCATTTGAAACAAGCGTATAATGTACCGATTAAATATTCTATGATTGAAGAACAGCAGGTTTTAGTC
CCCATCTTTACCATACAGTAA

SEQ ID NO:10 polypeptide sequence of Orf5

MNELSLDADKLLFGYDKPLYLPLTFQCKKGEVISVFGTNGKGKTTLLHSLAHVLPVMSGQIRQQGHIGFVPQ
SFSSPDYPVLEIVLMGRASKIGAFNLPSKTDETVALQMLACLDILHLAERNINMLSGGQRQLVLIARALATE
CQVLILDEPTAALDVYNQXRVLQLIRFLATEQKMTIIFSTHDPYHSLCVADNVLLLLPNQQWKYGIASQILT
ESHLKQAYNVPIKYSMIEEQQVLVPIFTIQ.

SEQ ID NO:11 polynucleotide sequence of Orf6

ATGAAGTCTATGTTAGCAAATCAGCGAGGTTTTATAACATCGCTGATTTTTATCTTGTTTATCATCGTAT
TGTTCACTTTAAATATTGGCACTTTTTCGTTATCAACCGGAAAAGTGATGTCCATTTTATCTAAGCCTTT
TCTTTCGCAACACGCGTCTTTTACACCTATGGAATACCATATTGTTTGGCATGTACGCTTACCACGCATC
ATTATGGCATTTTTTTCAGGGGGGATCTGAGCGATGAGTGGTGCAACACTACAGGGCGTTTTTCATAATC
CCCTTGTTGATCCTCATATTATTGGTGTCACATCAGGGGCAGTTTTTGGAGGCAGTTTAGCAATTTTATT
AGGATTCCCATCTTATTTATTGATTCTATCCACATTTTCTTTTGGTTTATTGACATTATTCTTGATCTAT
GTAACCACAATGTTCATCGGAAAAGGCAATCGTATTGTATTAGTTTTAGCGGGTGTCATTTTAAGTGGTT
TCTTTAGCACTCTAGTGAGCTTAATCCAATATTTAGCGGATGCAGAAGAAGTTCTGCCGAGCATTGTATT
TTGGTTATTAGGAAGTTTTGCCACCACTAGTTGGGCAAAACTAGCTATATTGTTACCCTGCGTTTTTATT
GCAGCTTATTTATTATTCCGTTTACGGTGGCATATTAATGTGTTATCGCTAGGTGATATGCAAGCAAAAA
TGTTAGGCGTTTCCATTAAGAAAATGCGTTGGTTTGTTTTGCTACTTTGTGCATTGCTTGTAGCAACACA
AGTCGCTGTTAGTGGGAGTATTGGGTGGATAGGGCTTGTTATTCCTCATTTGACACGTTTTTTTGTAGGA
AGTGATCACCGTTATCTATTGCCCGCCTCCTTTTTGATTGGTGGGATTTTCATGATTGTTATTGATACAC
TTGCACGTACGTTAACTTCTGCAGAAATTCCTGTAGGTATTATCACCGCTCTTTTAGGAGCACCCATTTT
TACCTTGCTCCTATTAAAAACTTATCGAAAGAAGTCATTATGA

SEQ ID NO:12 polypeptide sequence of Orf6

MKSMLANQRGFITSLIFILFIIVLFTLNIGTFSLSTGKVMSILSKPFLSQHASFTPMEYHIVWHVRLPRIIM
AFFSGGIXAMSGATLQGVFHNPLVDPHIIGVTSGAVFGGSLAILLGFPSYLLILSTFSFGLLTLFLIYVTTM
FIGKGNRIVLVLAGVILSGFFSTLVSLIQYLADAEEVLPSIVFWLLGSFATTSWAKLAILLPCVFIAAYLLF
RLRWHINVLSLGDMQAKMLGVSIKKMRWFVLLLCALLVATQVAVSGSIGWIGLVIPHLTRFFVGSDHRYLLP
ASFLIGGIFMIVIDTLARTLTSAEIPVGIITALLGAPIFTLLLLKTYRKKSL.

SEQ ID NO:13 polynucleotide sequence of Orf7

ATGATTCAACGCTACGTTAAAATAGTCAGTATTGCTTTATTACTTTTCTTAGGTTCTATTAATAATGCGT
TTGCAGCACGTGTTATTACTGATCAATTAGGACGAAAGGTCACTATCCCAGATGAAGTTAATCGTGTTGT
TGTCTGACAGCATCAGACTTTAAATCTCCTTGCCCAGCTTGATGCAAAGGAAAGTGTAGTCGGAGTGTTA
TCAAGTTGGAAAAAACAATTAGGGAAAAACTATGCACCAAAAGAAATGATTGAGCAAATCGAACAGGCTG
GTGTGCCTGTTGTAGCCATTTCTTTGCGTGAAGATAAAAAAGGTGAAGAAGGAAAAGTCAACCCAGAAAT
GGAAGATGAAGAAGTTGCCTATAATAATGGTTTGAAACAAGGCATTTATTTAATTGGTGAAGTAATTAAT
CGACAAGCGCAAGCCCAAAAGCTAGTTACTTACACTTTTGAACAGCGTGAATTAGTGAGTCAACGTTTAA
GTAAGGTGCCTGATGAGCAGCGTGTTAGGGTCTATATTGCAAATCCAGATTTAGCGACTTATGGTTCTGG
AAAATATACAGGGTTAATGATGCTTCATGCTGGAGCGAAGAATGTGGCAGCTGAAACAATAAAAGGTTTT
AAACAAGTTTCGATTGAGCAAGTGATTCATTGGAATCCTGCAGTTATCTTCGTACAGGAACGTTATCCTC
AGGTTATCGAGCAAATTAAAAAGGATCCCTCTTGGCAAATTATTGATGCGGTGAAAAATCAACGTATCTA
TTTAATGCCGGAATATGCAAAAGCGTGGGGATATCCAATGCCTGAAGCATTAGCGATTGGTGAATTATGG
TTAGCAAAACAACTTTACCCTGAATTGTTTGCAGATGTTGATTTAGAGGAAAAAGTAAACCAATACTATA
AATTGTTCTATCGTATGCCATATAACCAGTAA

SEQ ID NO:14 polypeptide sequence of Orf7

MIQRYVKIVSIALLLFLGSINNAFAARVITDQLGRKVTIPDEVNRVVVXQHQTLNLLAQLDAKESVVGVLSS
WKKQLGKNYAPKEMIEQIEQAGVPVVAISLREDKKGEEGKVNPEMEDEEVAYNNGLKQGIYLIGEVINRQAQ

AQKLVTYTFEQRELVSQRLSKVPDEQRVRVYIANPDLATYGSGKYTGLMMLHAGAKNVAAETIKGFKQVSIE QVIHWNPAVIFVQERYPQVIEQIKKDPSWQIIDAVKNQRIYLMPEYAKAWGYPMPEALAIGELWLAKQLYPE LFADVDLEEKVNQYYKLFYRMPYNQ.

SEQ ID NO:15 polynucleotide sequence of Orf8

TTAAGCAAGCAAAATAGTTTAATCCGCCTTTCTTTAATTAGTCTACTTATTTCCACTTCTTTTTATTCTG TTCAATCTTTTGTGGCAGATAGTTCTGATAAAACTTGGCAGTTACAAACAGGCCAAGGTTTAGATGCTAA AATAGGTCAAGTGAATAATCAATTTACACAAGTTGATACCCGTTTAAATCGAACAGATTTACGTATTAAC CGCCTTGGCGCAAGTGCTGCGGCGTTGGCTTCATTAAAACCTGCACAATTAGGCGAAGATGATAAATTTG CATTATCTTTGGGCGTTGGTAGTTATAAAAATGCGCAGGCGATGGCAATGGGGGCTGTGTTTAAGCCAGC TGAAAACGTATTGCTTAATGTAGCGGGGAGTTTTTCTGGTTCGGAAAAAACCTTTGGCGCAGGTGTTTCT TGGAAATTCGGCAGCAAATCCAAACCTGCGGTTTCAACACAAAGTGCGGTCAATTCTGCGGAAGTTTTGC AACTGCGACAAGAAATATCGGCAATGCAAAAAGAATTGGCTGAATTGAAAAAAGCATTAAGAAAATAA

SEQ ID NO:16 polypeptide sequence of Orf8

LSKQNSLIRLSLISLLISTSFYSVQSFVADSSDKTWQLQTGQGLDAKIGQVNNQFTQVDTRLNRTDLRINRL GASAAALASLKPAQLGEDDKFALSLGVGSYKNAQAMAMGAVFKPAENVLLNVAGSFSGSEKTFGAGVSWKFG SKSKPAVSTQSAVNSAEVLQLRQEISAMQKELAELKKALRK.

SEQ ID NO:17 polynucleotide sequence of Orf9

ATGGAGCATTCTGTTCATAACAAACTGGTTTCTTTTATTTGGAGTATTGCAGACGATTGTCTGCGCGATG TGTATGTGCGCGGTAAATATCGTGATGTGATTTTACCGATGTTTGTGCTTCGTCGTTTGGATACTTTACT TGAGCCAAGCAAAGATGCCGTATTGGAAGAAATGCGTTTTCAAAAAGAAGAATTGGCATTCACCGAATTG GATGACCTTCCCCTTAAAAAAATTACCGGTCATGTTTTTTATAACACCTCAAAATGGACATTAAAATCCC TCTATCAAACCGCCAGCAATACGCCGCAGTATATGCTGGCCAATTTTGAAGAATATCTTGATGGTTTCAG CACCAACATTCATGAAATCATCAACTGCTTCAAGCTGCGTGAACAAATCCGCCATATGTCCCATAAAAAT GTTTTGCTGAGCGTGTTGGAAAAATTTGTATCGCCCTATATCAATCTTACCCCTAAAGAACAACAAGACC CTGAGGGCAACAAATTACCAGCGCTGACCAATCTGGGCATGGGCTATGTATTTGAAGAACTGATTCGTAA ATTTAACGAAGAAAATAACGAAGAAGCTGGCGAACACTTTACCCCACGCGAAGTGATCGAGCTGATGACG CATTTAGTCTTTGATCCGCTCAAAGACCAAATTCCGGCCATTATTACGATTTACGACCCAGCTTGCGGCA GCGGTGGCATGCTGACCGAGTCGCAAAACTTTATTGAGCAAAAAATATCCGCTATCTGAATCACAAGGCGA GCGTTCCATCTTTTTGTTTGGTAAAGAAACCAATGATGAAACCTATGCCATTTGTAAATCTGACATGATG ATTAAAGGTGATAATCCCGAAAACATCAAAGTCGGCTCAACCCTTGCTACAGATAGCTTCCAAGGTAATC ACTTTGACTTTATGCTTTCCAACCCGCCATATGGCAAAAGCTGGAGCAAAGATCAAGCCTATATCAAAGA CGGCAATGAGGTTATCGACAGTCGCTTTAAAGTTACCTTACCAGATTACTGGGGCAATGTAGAAACCCTT GATGCTACCCCACGCTCCAGCGATGGACAGCTGCTATTCCTAATGGAAATGGTCAGCAAAATGAAATCGC CGAATGACAACAAAATCGGCAGCCGAGTGGCCTCCGTGCATAACGGCTCAAGCCTGTTTACCGGCGATGC AGGTTCAGGAGAAAGCAACATTCGTCGCCATATTATTGAAAAAGATTTGCTCGAAGCCATCGTACAGCTG CCTAACAACCTGTTTTATAACACAGGTATTACCACTTATATTTGGTTGCTGTCCAACAACAAACCTGAAG CACGCAAAGGCAAAGTTCAGCTCATTGATGCCAGCCTCTTATTCCGCAAATTGCGTAAAAACCTTGGCGA TAAAAACTGCGAATTTGTACCTGAACATATCGCCGAAATTACCCAAAACTATCTTGATTTCACTGCCAAA GCGCGCGAAACCGACAGCCAAAATGAAGCAGTCGGCCTGGCTTCGCAGATTTTTGACAATCAAGATTTCG GCTATTACAAAGTCACCATCGAACGCCCGGATCGCCGTTCTGCCCAATTTACCGCCGAAAATATCTCGCC TTTACGGTTTGACAAGGCTTTGTTTGAGCCGATGCAATATCTTTATCGGCAATATGGCGAACAAATTTAC

AACGCCGGATTTTTAGCCCAAACCGAGCAAGAAATTACCGCTTGGTGCGAAGCGCAGGGCATAGCCTTAA
ACAACAAAAACAAGACCAAGCTGCTGGACGTCAAAACCTGGGAAAAAGCCGCCGCACTTTTTCAGACGGC
ATCAACCTTGCTCGAACATTTCGGCGAACAACAATTTGACGATTTCAACCAATTCAAACAAGCCGTGGAA
TGCCGTCTGAAAGCCGAAAAAATCCCCCTTTCTGCCACAGAGAAAAAGGCCGTTTTCAATGCCGTAAGTT
GGTACGACGAAAATTCAGCCAAAGTGATTGCCAAAACACTCAAGCTCAAACCAAACGAATTGGACGCCCT
TTGCCAACGCTACCAATGCCAAGCCGACGAGCTGGCAGACTTTGGCTATTACGCCACCGGCAAAGCAGGC
GAATATATCCTATATGAAACGAGCAGCGACTTGCGCGACAGCGAATCCATACCGCTCAAACAAAATATCC
ACGACTATTTCAAAGCCGAAGTGCAAGCGCACATCAGCGAAGCATGGCTGAATATGGAAAGCGTAAAAAT
CGGCTATGAAATCAGCTTCAACAAATACTTCTACCGCCACAAACCATTACGCAGCCTTGCAGAAGTTGCCCA
AGATATTTTGGCGTTAGAAAAACAGGCTGACGGCTTGATTAGTGAAATTCTAGAGGCTTAA

SEQ ID NO:18 polypeptide sequence of Orf9

MEHSVHNKLVSFIWSIADDCLRDVYVRGKYRDVILPMFVLRRLDTLLEPSKDAVLEEMRFQKEELAFTELDD
LPLKKITGHVFYNTSKWTLKSLYQTASNTPQYMLANFEEYLDGFSTNIHEIINCFKLREQIRHMSHKNVLLS
VLEKFVSPYINLTPKEQQDPEGNKLPALTNLGMGYVFEELIRKFNEENNEEAGEHFTPREVIELMTHLVFDP
LKDQIPAIITIYDPACGSGGMLTESQNFIEQKYPLSESQGERSIFLFGKETNDETYAICKSDMMIKGDNPEN
IKVGSTLATDSFQGNHFDFMLSNPPYGKSWSKDQAYIKDGNEVIDSRFKVTLPDYWGNVETLDATPRSSDGQ
LLFLMEMVSKMKSPNDNKIGSRVASVHNGSSLFTGDAGSGESNIRRHIIEKDLLEAIVQLPNNLFYNTGITT
YIWLLSNNKPEARKGKVQLIDASLLFRKLRKNLGDKNCEFVPEHIAEITQNYLDFTAKARETDSQNEAVGLA
SQIFDNQDFGYYKVTIERPDRRSAQFTAENISPLRFDKALFEPMQYLYRQYGEQIYNAGFLAQTEQEITAWC
EAQGIALNNKNKTKLLDVKTWEKAAALFQTASTLLEHFGEQQFDDFNQFKQAVECRLKAEKIPLSATEKKAV
FNAVSWYDENSAKVIAKTLKLKPNELDALCQRYQCQADELADFGYYATGKAGEYILYETSSDLRDSESIPLK
QNIHDYFKAEVQAHISEAWLNMESVKIGYEISFNKYFYRHKPLRSLAEVAQDILALEKQADGLISEILEA.

SEQ ID NO:19 polynucleotide sequence of Orf10

ATGCAGCCGGAAAACCAATATTTTGAGCGCAAAGGACTAGGAGAAAAAGACATCAAGCCAACTAAAATAG
CTGAAGAATTAGTTGGAATGCTCAATGCTGATGGCGGAGTTTTGGCTTTTGGTGTGGCAGATAATGGCGA
AATCCAAGACTTGAATAGCCTTGGCGATAAATTAGATGATTATCGGAAATTGGTTTTCGATTTTATTGCA
CCGCCTTGTCGGATTGGACTGGAAGAAATTCTGGTTGATGGAAAATTAGTTTTCTTATTCCACGTAGAGC
AAGATTTAGAGCGTATTTATTGTCGCAAAGACAATGAAAATGTGTTCTTACGTGTAGCAGATAGTAATCG
AGGCCCTCTCACCAGAGAACAAATCAAAAATCTTGAATATGATAAAAATATCCGTCTATTTGAAGATGAA
ATAGTTCCTGATTTTAATGAAGAAGATTTAGATCAAGAATTATTAGAGCTATATAAAAAGAAAGTTAATT
TTACCTCCGATAATATCTTAGATTTATTATACAAGCGAAATTTATTAACCAAAAAGGAAGGTTGTTATCA
GTTTAAAAAATCAGCCATTTTACTCTTTTCTACCATGCCGGAACGTTACATTCCTTCAGCATCAGTCCGC
TATGTTCGTTATGAAGGTACAGTAGCGAAAGTCGGTACTGAGCATAATGTGATAAAAGACCAACGTTTTG
AAAATAATATTCCAAAGCTAATTGAGGAGCTGACCTATTTTTAAGAGCCTCTTTAAGGGATTATTACTT
TCTTGATGTCAATCAGGGAAAATTTATCAAAGTACCGGAATATCCTGA

SEQ ID NO:20 polypeptide sequence of Orf10

MQPENQYFERKGLGEKDIKPTKIAEELVGMLNADGGVLAFGVADNGEIQDLNSLGDKLDDYRKLVFDFIAPP
CRIGLEEILVDGKLVFLFHVEQDLERIYCRKDNENVFLRVADSNRGPLTREQIKNLEYDKNIRLFEDEIVPD
FNEEDLDQELLELYKKKVNFTSDNILDLLYKRNLLTKKEGCYQFKKSAILLFSTMPERYIPSASVRYVRYEG
TVAKVGTEHNVIKDQRFENNIPKLIEELTYFLRASLRDYYFLDVNQGKFIKVPEYP

SEQ ID NO:21 polynucleotide sequence of Orf11

ATGTCAATCAGGGAAAATTTATCAAAGTACCCGGAATATCCTGAAGAAGCTTGGTTAGAAGGTGTTGTAA
ATGCGCTTTGTCATCGTTCTTACAATGTTCAAGGTAATGTTATTTATATTAAACATTTCGACGATCGTCT
TGAAATTAGTAATAGTGGCCCTCTCCCTGCTCAAGTCACCATTGAAAATATTAAAACGGAACGATTCGCT

CGGAATCCACGTATAGCACGAGTTTTAGAGGATCTTGGGTATGTCCGTCAGCTTAATGAAGGCGTTTCCC
GTATTTATGAGTCAATGGAAAAATCATTATTGGCAAAGCCTGAATATAGAGAACAAAACAACAATGTTTA
TCTAACATTGCGCAACCGTGTTACCGCACATGAAAAAACGGTATCTACAGCCACTATGCTGCAGATTGAA
AAAGAATGGACAAACTACAACGACACCCAAAAAGCCATTTTGCTTTATCTATTTACAAATGGTACGGCGA
TATTGTCAGAATTAGTTGACTATACAAAAATCAATCAGAATTCGATCCGAGCGTATTTAAATGCCTTTAT
TCAGCAAGGTATTATTGAAAGACAAAGTGTAAAACAGCGTGACCCCAATGCCAAATATGCTTTTAGAAAA
GATTAA

SEQ ID NO:22 polypeptide sequence of Orf11

MSIRENLSKYPEYPEEAWLEGVVNALCHRSYNVQGNVIYIKHFDDRLEISNSGPLPAQVTIENIKTERFARN
PRIARVLEDLGYVRQLNEGVSRIYESMEKSLLAKPEYREQNNNVYLTLRNRVTAHEKTVSTATMLQIEKEWT
NYNDTQKAILLYLFTNGTAILSELVDYTKINQNSIRAYLNAFIQQGIIERQSVKQRDPNAKYAFRKD.

SEQ ID NO:23 polynucleotide sequence of Orf12

TTGCAAATGAGACGATACGAGCGTTACAAAGATTCAGGTGTGGATTGGCTAGGGGAGGTACCGAGCCATT
GGGAGTTAAAACGCTTGAAACAATTATTTGTTGAAAAAAAACATAAGCAAAGCCTGTCTCTTAATTGTGG
AGCCATTAGTTTTGGTAAAGTTATTGAAAAATCGGATGATAAAGTAACAGAGGCAACAAAACGTTCATAT
CAAGAGGTGTTAAAAGGCGAGTTTTTAATAAATCCTTTAAACTTAAATTATGACCTAATTAGTTTGAGAA
TTGCTTTATCAGAAATAGACGTTGTTGTAAGTGCCGGTTACATTGTTTTAAAAGAAAAACAAATAATTAA
TAAAAAATACTTTTCGTATTTATTACATAGATACGATGTTGCATATATGAAATTATTAGGTTCAGGTGTA
AGACAAACGATTAACTATGGGCATATTTCAGACAGTATTTTGGTTATTCCACCTCTCTCCGAACAACAAA
AAATCGCGCAATTCCTAGACGATAAAACCGCTAAAATCGATCAGGCGGTGGATTTGGCGGAAAAGCAGAT
TGCCCTGTTGAAAGAGCACAAGCAGATCCTGATTCAAAATGCCGTAACCCGAGGCTTAAACCCTGATGTG
CCGTTAAAAGATTCCGGCGTGGAATGGATAGGGCAAGTGCCGGAGCATTGGGATGTGCAACGTTCAAAAT
TCATTTTCAAGAAAATAGAAAGAAAAGTGAATGAGGAAGACCAAATTGTTACTTGTTTTAGGGATGGGCA
AGTAACTCTGAGAGCTAATCGAAGAACTGAAGGATTTACAAATGCGCTAAAAGAACACGGCTACCAAGGA
ATTAGAAAAGGTGATTTAGTTATTCACGCTATGGATGCTTTTGCAGGGGCAATTGGTATTTCTGATTCAG
ATGGTAAAGCAACACCAGTTTATTCCGTTTGTTTGCCTCATGATAAACAAAAAATCGATGTCTATTTTTA
CGCTTATTACTTAAGAAATCTTGCATTATCAGGATTTATTAGCTCCTTAGCTAAAGGAATTAGAGAGCGT
TCAACAGATTTTCGCTATTCTGATTTTGCAGAATTATTACTACCTATTCCTCCATATTTAGAACAGCAAA
AAATTGCCGACTACCTAGATAAACAAACCTCTAAAATTGATCGAGCAATCGCATTAAAAACAGCCCATAT
TGAAAAGCTGAAAGAATATAAAAGCGTGTTGATTAACGATGTGGTGACCGGCAAGGTGCGGGTATAG

SEQ ID NO:24 polypeptide sequence of Orf12

LQMRRYERYKDSGVDWLGEVPSHWELKRLKQLFVEKKHKQSLSLNCGAISFGKVIEKSDDKVTEATKRSYQE
VLKGEFLINPLNLNYDLISLRIALSEIDVVVSAGYIVLKEKQIINKKYFSYLLHRYDVAYMKLLGSGVRQTI
NYGHISDSILVIPPLSEQQKIAQFLDDKTAKIDQAVDLAEKQIALLKEHKQILIQNAVTRGLNPDVPLKDSG
VEWIGQVPEHWDVQRSKFIFKKIERKVNEEDQIVTCFRDGQVTLRANRRTEGFTNALKEHGYQGIRKGDLVI
HAMDAFAGAIGISDSDGKATPVYSVCLPHDKQKIDVYFYAYYLRNLALSGFISSLAKGIRERSTDFRYSDFA
ELLLPIPPYLEQQKIADYLDKQTSKIDRAIALKTAHIEKLKEYKSVLINDVVTGKVRV.

SEQ ID NO:25 polynucleotide sequence of Orf13

ATGGTTTCAGGAACTAAGGAAAAAGATTTAGAAATTGCCATCGAAAAAGCCTTAACTGGCACTTGGCGTG
AAAACATGGAAAATAAGCTGGGCGAGCCGAAGGCTGAATACCTGCCGCGCCATCATGGTTTTAAACTGGC
ATTTTCACAGGATTTTGATGCGCAGTTTGCCATCGACACACGTCTGTTTTGGCAATTCCTGCAAACCAGC
CAAGAGGCAGAACTTGCCCGTTTTCAACAACTCAACCCAAACGACTGGCAGCGTAAAATTTTGGAGCGAT
TAGACCGCCAAATAAAGAAAAACGGCGTGTTGCACCTGCTGAAAAAAGGCTTGGATATTGATAGCGCCCA

TTTTGATTTGCTCTACCCCGTTCCGCTTGCCAGCAGCGGCGAAAAGGTCAAGCAGCGTTTTGAACAGAAT
TTGTTTAGCTGTATGCGTCAAGTGCCTTATTCTGCCTCAAGCAATGAAACGGTGGATATGGTGCTGTTTG
CCAATGGCTTGCCGATTATTGCCCTTGAGCTGAAAAACCATTGGACAGGTCAGACAGCCATTGATGCGCA
AAAACAATACCTCAACCGTGATTTAAGCCAAACGTTGTTCCATTTCGGGCGTTGTTTGGCGCATTTTGCC
TTAGATACGGAAGAAGCTTATATGACCACCAAATTGGCGGGGCCTGCTACGTTTTTCTTGCCGTTTAACT
TGGGCAACAACTGCGGTAAGGGTAATCCGCCCAATCCCAATGGACACCGCACGGCGTATTTATGGCAAGA
GGTGTTCGGCAAAGCAAGCCTTGCCAACATTATTCAGCATTTTATGCGCTTAGACGGTTCAACCAAAGAT
CCGTTGGATAAACGTACCCTCTTTTTCCCTCGCTATCACCAATTAGATGTGGTCCGCCGTTTGATTGCTG
ATGTCAGTGAACATGGCGTGGGTAAACGTTATTTGATTCAACATTCTGCCGGTTCGGGCAAGTCTAATTC
CATTACTTGGCTGGCGTATCAGTTGATTGAGGCATATCCGCGCAATGAAAAGGCGGCAAACGGTAGAGAG
GCAGACCGCCCGATTTTTGATTCGGTGATTGTCGTAACCGACCGTCGTTTGTTGGATAAGCAACTGCGCG
ACAATATCAAAGATTTTTCAGAAGTTAAAAACATTGTTGCGCCGGCGTTGAGTTCGGCAGAGTTGCGCCA
ATCGCTTGAGCAGGGCAAAAAAATCATTATTACCACGATTCAAAAATTCCCGTTTATTGTCGATGGCATT
GCTGATTTAGGCGACAAACAATTTGCGGTGATTATTGATGAGGCACACAGCTCACAATCAGGTTCGGCAC
ACGACAATATGAACCGGGCCATCGGCAAAACGGAAGACCTTGATGCTGAAGATGTGCAAGATTTGATTTT
ACAAACCATGCAATCCCGCAAAATGCACGGCAATGCGTCGTATTTTGCTTTCACCGCCACACCGAAAAAC
AGCACTTTGGAAAAATTCGGCGAAAAACAGGCGGATGGCAAGTTTAAGCCGTTCCACCTTTATTCTATGA
AGCAGGCGATTGAAGAAGGCTTTATTTTGGATGTAATCGCCAATTACACCACCTATAAAAGTTTTTATGA
GATCACTAAGTCGATTGAAGATAATCCGGAGTTTGATAGTAAAAAGGCTCAAAGCCGTCTGAAAGCCTAT
GTGGAGCGTTCGCAACAAACGATTGATACTAAAGCGGAGATAATGCTGGATCATTTTATTTACCAAGTTT
TCAACCGTAAAAAACTCAAAGGCAAAGCCAAGGGAATGGTGGTAACGCAAAATATTGAAACCGCCATCCG
CTATTTTCAGGCGTTAAAACATTTGCTGGCCGGGCGGGGTAATCCGTTTAAAATTGCGATTGCGTTTTCA
GGCAGTAAAGTGGTTGACGGTGTCGAATACACCGAAGCGGAAATGAACGGCTTTGCAGAAAGCGAAACCA
AAGAGTATTTCGATCAAGATGAATATCGTTTGCTGGTGGTCGCCAATAAATATCTGACCGGTTTCGATCA
GCCGAAATTGTGTGCCATGTATGTGGATAAGAAACTCTCCGGCGTGCTTTGCGTGCAGGCTTTATCTCGT
TTGAATCGCAGTGCGAATAAGTTGAGTAAACGCACGGAAGATTTGTTTGTATTGGACTTTTTTAACAGCG
TTGAAGATATTCAGCAGGCATTTGAGCCGTTTTATACTTCTACTTCGTTGTCGCAGGCAACCGATGTCAA
TGTCTTGCATGATTTGAAAGACCGGTTGGATGAAACCGGCGTGTACGAACAAGCGGAGGTCAACGATTTT
ACTGAAGGCTATTTTGCCAATAAAGACGCACAGCAATTAAGCAGTATGATTGATGTGGCTGTCCAACGTT
TTGATGATGAATTGGAATTGGATTTGGATCGAAATGAAAAAGTTGATTTTAAAATCAAGGCAAAACAGTT
TTTAAAAATTTACGGGCAAATGGCCTCCATCATCAATTTTGAAAATATCGCTTGGGAAAAGCTCTATTGG
TTCCTCAAATTCTTAGTACCCAAATTAAAAGTACAAGACCCGATGGATGAATTTGATGAAATTTTAGATG
CAGTGGATTTAAGCTCTTACGGCTTGGCGCACACCAAGCTGAATTACAGCATTAAATTAGATGATGAAGA
AACAGAGCTTGACCCGCAAAACCCCAATCCGCGCGGTACGCATGGTGAAGATAAAGAAAAAGATCCGATT
GATGAAATTATTCGTGTATTTAACGAAAGATGGTTTCAAGATTGGAGCGCAACGCCGGATGAGCAACGGG
TAAAATTTATCAATATTACCGAGCGCATCCGCAGCCATAAAGACTTTGAGCAGAAATATCAAAATAACCC
GGATATTCATACCCGTGAATTGGCTTTCCAAGCCATTTTGCGCGATGTGATGAGCGAACGCCATAGGGAT
GAATTAGAGCTATACAAACTTTTTGCCAAAGATGCCGCATTTAGAACCGCTTGGACGCAAAGTTTGCAAC
GGGCTTTGGCTGGATAG

SEQ ID NO:26 polypeptide sequence of Orf13

MVSGTKEKDLEIAIEKALTGTWRENMENKLGEPKAEYLPRHHGFKLAFSQDFDAQFAIDTRLFWQFLQTSQE
AELARFQQLNPNDWQRKILERLDRQIKKNGVLHLLKKGLDIDSAHFDLLYPVPLASSGEKVKQRFEQNLFSC
MRQVPYSASSNETVDMVLFANGLPIIALELKNHWTGQTAIDAQKQYLNRDLSQTLFHFGRCLAHFALDTEEA

YMTTKLAGPATFFLPFNLGNNCGKGNPPNPNGHRTAYLWQEVFGKASLANIIQHFMRLDGSTKDPLDKRTLF
FPRYHQLDVVRRLIADVSEHGVGKRYLIQHSAGSGKSNSITWLAYQLIEAYPRNEKAANGREADRPIFDSVI
VVTDRRLLDKQLRDNIKDFSEVKNIVAPALSSAELRQSLEQGKKIIITTIQKFPFIVDGIADLGDKQFAVII
DEAHSSQSGSAHDNMNRAIGKTEDLDAEDVQDLILQTMQSRKMHGNASYFAFTATPKNSTLEKFGEKQADGK
FKPFHLYSMKQAIEEGFILDVIANYTTYKSFYEITKSIEDNPEFDSKKAQSRLKAYVERSQQTIDTKAEIML
DHFIYQVFNRKKLKGKAKGMVVTQNIETAIRYFQALKHLLAGRGNPFKIAIAFSGSKVVDGVEYTEAEMNGF
AESETKEYFDQDEYRLLVVANKYLTGFDQPKLCAMYVDKKLSGVLCVQALSRLNRSANKLSKRTEDLFVLDF

FNSVEDIQQAFEPFYTSTSLSQATDVNVLHDLKDRLDETGVYEQAEVNDFTEGYFANKDAQQLSSMIDVAVQ
RFDDELELDLDRNEKVDFKIKAKQFLKIYGQMASIINFENIAWEKLYWFLKFLVPKLKVQDPMDEFDEILDA
VDLSSYGLAHTKLNYSIKLDDEETELDPQNPNPRGTHGEDKEKDPIDEIIRVFNERWFQDWSATPDEQRVKF
INITERIRSHKDFEQKYQNNPDIHTRELAFQAILRDVMSERHRDELELYKLFAKDAAFRTAWTQSLQRALAG

.

SEQ ID NO:27 polynucleotide sequence of Orf14

ATGTCTGAATATAAATTAAACCCACCGACAGTGTCTTCTTATACTGAAAATATGATGCTTAAAGTTTTAT
TTGAGCATAAAGGTTTTTCCGAAGTGTTTCGGGAGACTAGCTGGCGAAGTGATGAAATTGCCAGTGCATT
TGGGCTGCCTGAAGAATTAGAGAATGATAAAAATTTACGCACGGTTGCTCGTCGGCTTTTAAAAGAGCGG
TATAAAAAACTCCAAAAATCCACCGCACTTTTACCTGAGTTATGGAAACAGGCGTATGAAAATTTGGCAA
CGTTGGCAGAATTTTTGCAACTGAATCCCGTTGAACAGGAACTTCTCCGCTTTGCCATGCATTTACGTAG
TGAAGGAGCTATGCGAGATTTGTTTGGCTACTTGCCGAAATCGGATTTACAAAGAACGGCTGCGATCATG
GCGGATTTACTTAAACAGCCGAAAAATCAGATTCTATCTGCCTTAAAGAAAGGCAGTAAACTCGATGCTT
ATGGCCTGATTGATCGCGATTATCGCCCCGATAGTGTGCATGATTATTTAGATTGGGGCGAAACCTTAGA
TTTTGATGAATTTGTGACACAACCATTAAACGAAAACGTCCTATTAAAATCTTGTACGGAAGTCGCTCAA
GTGCCAAGTCTGCAACTGGATGATTTTGACCATATTGCCGGCATGAAAGAGATGATGTTGACTTATTTGC
AACAAGCACTAAAACATCATCGAAAAGGCGTGAATCTTTTAATTTATGGCGTGCCTGGCACTGGTAAAAC
AGAATTCGCCGGGTTGCTTGCACAGGCGTTGGGATTTCGGCGTATAACATTACTTACATGGATTCTGAC
GGAGATGTTGTGGAGGCAGAGCAACGCCTGAACTACAGTCGTCTTGCTCAAACGCTATTGAACGGCAAGC
AGGCGCTTTTAATTTTTGATGAAATTGAAGATGTGTTTAACGGCTCGTTTATGGAGCGTTCTGTTGCACA
AAAAAATAAAGCGTGGACAAATCAGTTATTGGAAAACAATAACGTGCCGATGATTTGGTTATCTAACTCT
GTTTCGGGCATAGATCCTGCTTTTTTACGCCGCTTTGATTTTATTTTAGAAATGCCAGATTTGCCGTTGA
AAAATAAGTCAGCACTGATTACGCAACTGACTGAGGGAAAATTAAGTCCGGCCTATGTGCAGCATTTTGC
TAAAGTGCGGTCATTAACGCCGGCGATTTTAAGCCGCACAATTCGGGTGGCAAAGGAACTCAATACATCA
AATTTTGCTGAGACTTTGCTCATGATGTTTAATCAAACGTTAAAATCGCAAAATAAACCGAAAATTGAAC
CGCTTGTTTTAGGCAAAGCCGACTACAACTTGGATTATGTGGCTTGTAACGACAATATTCATCGTATTAG
TGAAGGGTTAAAACGGTCGAAAAAAGGGCGAATTTGTTGCTATGGCCCGCCGGGAACAGGAAAAACTGCT
TGGGCAGCGTGGCTTGCGGAACAGTTGGACATGCCGCTATTGCTAAGACAAGGCTCAGATTTACTTAATC
CTTATGTGGGCGGACAGAACAAAATATTGCTCAAGCCTTTGAACAAGCGAAAGCCGATAATGCAATATT
GGTGCTAGATGAAGTAGATACGTTCTTATTTTCTAGAGAAGGCGCAAATCGAAGCTGGGAGCGTTCGCAA
GTGAATGAAATGCTAACACAAATTGAACGCTTTGAGGGCCTGATGGTGGTATCAACAAATTTAATTGAGG
TTCTTGATCACGCAGCTTTACGCCGTTTTGATTTAAAATTGAAGTTTGATTATTTAACGCTCAAACAACG
CTTAGATTTTGCTAAACAACAAGCAGAAATTTTAGGATTGCCGTTGTTATCGGAAGAGGATTTAAGTCAG
ATTGAATCGCTTAATCTGCTGACACCAGGGGATTTTGCTGCAGTGGCTCGTCGTCACCAATTTTCCCCTT
TTCACAAGGTGCAAGATTGGCTGATGGCACTACAAGGGGAATGTGAAGTGAAACCAGCGTTTTCTGCAAC
GACAAGGCGGATAGGGTTCTAA

SEQ ID NO:28 polypeptide sequence of Orf14

MSEYKLNPPTVSSYTENMMLKVLFEHKGFSEVFRETSWRSDEIASAFGLPEELENDKNLRTVARRLLKERYK
KLQKSTALLPELWKQAYENLATLAEFLQLNPVEQELLRFAMHLRSEGAMRDLFGYLPKSDLQRTAAIMADLL
KQPKNQILSALKKGSKLDAYGLIDRDYRPDSVHDYLDWGETLDFDEFVTQPLNENVLLKSCTEVAQVPSLQL
DDFDHIAGMKEMMLTYLQQALKHHRKGVNLLIYGVPGTGKTEFAGLLAQALGISAYNITYMDSDGDVVEAEQ
RLNYSRLAQTLLNGKQALLIFDEIEDVFNGSFMERSVAQKNKAWTNQLLENNNVPMIWLSNSVSGIDPAFLR
RFDFILEMPDLPLKNKSALITQLTEGKLSPAYVQHFAKVRSLTPAILSRTIRVAKELNTSNFAETLLMMFNQ
TLKSQNKPKIEPLVLGKADYNLDYVACNDNIHRISEGLKRSKKGRICCYGPPGTGKTAWAAWLAEQLDMPLL
LRQGSDLLNPYVGGTEQNIAQAFEQAKADNAILVLDEVDTFLFSREGANRSWERSQVNEMLTQIERFEGLMV
VSTNLIEVLDHAALRRFDLKLKFDYLTLKQRLDFAKQQAEILGLPLLSEEDLSQIESLNLLTPGDFAAVARR
HQFSPFHKVQDWLMALQGECEVKPAFSATTRRIGF.

SEQ ID NO:29 polynucleotide sequence of Orf15

ATGTTTGAAAAAATTGAACCTACTAATATTCGTTTTATTAAATTAGGCATAAAAGGATGTTGGGAAAAAG
ATTGTATTGATAAAAATAGTACAGCAAGTACAAAAAATACGATTCGTCTTGGCTATGAATCTACATCAGA
GATTCACAAAGAATGTTTGAATAATCAATGGGATAGTTGTATTGAATATTGTAAAACTTATTGGAGTGAC
CATACAGGAACTGTTTCAAATCACTTGAGACAAATTCAAGATTTTTATCAACTTGGGGAAGATACACTTT
GGATCACCTTCTTTGGACGTAAATTATATTGGGCTTTTTGCAGTAAAGAGGTTGTTGAGGAAAGCGATGG
TTCTAGAACAAGAAAAGTTATTAGTAACAATGGGAATTGGTCTTGCGTTGATGCTAACGGTAAAGAGCTT
TTAGTCGATAATCTTGATGGTAGAGTAACAAAGGTCCAAGCCTATAGAGGGACGATTTGTGGTGTTGAGA
TGGAGGACTATTTAATACGTCGTATAAATGGTGAAGTTATTGAGGAAATTACAGAAGCGAAAGAGGCGTA
TGAAACATTAATTAAATCAGTTGAAAAATTAATTAAAGGTTTATGGTGGAGTGACTTTGAACTTTTAACG
GATCTTGTTTTTTCTAAATTAGGATGGCAACGATACTCTGTTTTAGGTAAAACGGAGAAAGGAATAGATC
TTGATTTGTATTCGTCTTCAACGCAGAAGAGAGTATTTGTGCAAATTAAGTCAGATACGGATATTAAACA
ATTAGACGAATATGTTTCGAACTTTGAAAGTGAATATAAAAACTATGGTTATTCAGAAATGTATTACGTA
TATCATTCTGGTTTAGAAAACATAGATGAAAAACAATATCAAGCTAAAGGAATTAAGCTTGTAAATGGCC
GAAAAATGGCAGAGCTTGTAATTAGTGCTGGTTTAGTTGAATGGTTGATTAACAAACGTTCTTAA

SEQ ID NO:30 polypeptide sequence of Orf15

MFEKIEPTNIRFIKLGIKGCWEKDCIDKNSTASTKNTIRLGYESTSEIHKECLNNQWDSCIEYCKTYWSDHT
GTVSNHLRQIQDFYQLGEDTLWITFFGRKLYWAFCSKEVVEESDGSRTRKVISNNGNWSCVDANGKELLVDN
LDGRVTKVQAYRGTICGVEMEDYLIRRINGEVIEEITEAKEAYETLIKSVEKLIKGLWWSDFELLTDLVFSK
LGWQRYSVLGKTEKGIDLDLYSSSTQKRVFVQIKSDTDIKQLDEYVSNFESEYKNYGYSEMYYVYHSGLENI
DEKQYQAKGIKLVNGRKMAELVISAGLVEWLINKRS.

SEQ ID NO:31 polynucleotide sequence of Orf16

TTACCCTTTGCCAACAAAATTGGCAGCAACAAGCGACGCAACCAAGATGCCCTTTTTAATGGCGAGGCGG
TGTTTCAATATAAACTCAAAACGGCTGAAAAACGCCTTGAAAACCGACCGCACTTTATTGTGGGCGTGGC
AGATGGTATTTCTAATAGCAACCGACCTGAAAAAGCGAGCAAATTGGCTATGCAATTATTAAGCCAAATG
GAAAGTATAAACCGTCAAACGATCTACGATTTACAATCCAGTTTATCAGCAGAATTAGCTGAGGATTATT
TTGGTTCGGCGACCACATTTGTGGCTGCCGAAATTGATCAAATAACCCGTAAAGCGAAAATTCTCAGCGT
AGGCGATAGTCGTGCTTATTTAATTGATGCCCAAGGAAAATGGCAACAAATCACCCAAGATCATTCTATT
CTTTCTGAATTATTGACTGATTTCCCCGATAAAAAAGAAGAAGATTTTGCCACGATTTATGGCGGCGTTT
CTTCTTGTTTAGTCGCCGATTATTCCGAATTTCAAGATAAAATTTTTTATCAAGAAATTGAAATTCAGCA
AGGGGAAAGTTTATTACTTTGTTCTGACGGCTTGACCGACGGGCTTTCAGATGAAATGCGCGAAAAAATT
TGGCAGAAATATCCCGATGATAAATATCGCCTTACGGTTTGCCGCAAGATGATTGAGAAGCAATCGTTTT
CGGATGATTTGTCGGTAGTTTGTTGTCATTCTATTATTGAGTAA

SEQ ID NO:32 polypeptide sequence of Orf16

LPFANKIGSNKRRNQDALFNGEAVFQYKLKTAEKRLENRPHFIVGVADGISNSNRPEKASKLAMQLLSQMES
INRQTIYDLQSSLSAELAEDYFGSATTFVAAEIDQITRKAKILSVGDSRAYLIDAQGKWQQITQDHSILSEL

LTDFPDKKEEDFATIYGGVSSCLVADYSEFQDKIFYQEIEIQQGESLLLCSDGLTDGLSDEMREKIWQKYPD
DKYRLTVCRKMIEKQSFSDDLSVVCCHSIIE

SEQ ID NO:33 polynucleotide sequence of Orf17

ATGAAAAATGATTTGAATTATGCAGTGGAACTTATCCGCAAAGCGGATGGCATTTTAATTACAGCTGGTG
CGGGTATGAGCGTGGATTCTGGGCTTCCCGATTTCCGCAGCGTTGGCGGATTTTGGAATGCTTATCCTAT
GTTTAAAGAACATAATATATCTTTTGAAGAGATCGCAACGCCACTAGCTTATAAGCATAATCAGGAACTA
GCCTATTGGTTTTATGGGCATCGATTAGTTCAATACCGAAATACTCTTCCTCACGAAGGGTATCAGATTT
TAAAATGCTGGGCGGGAGATAAACCTCATGGATATTTTGTTTTTACCAGTAATGTTGATGGGCATTTTCA
AAAGGCTGGTTTTAATGATAGCCATGTTTATGAAGTACATGGTACTTTGGAGCGTCTTCAATGTGTCAAT
AATTGTCGAGGATTAAGTTGGTCTGCATCAAGTTTTCAACCTGTCGTGGATAATGAAAACTTATGTTTAA
CCAGTGAAAAACCACATTTGCCTTATTGTGGGGGCTTTGCTCGTCAAAATGTACTAATGTTTAATGATTG
GAGTTATGCAAGTCAATATCAGGATTTTAAAAAAGTGCGGTTAGAATCGTGGTTAAAAGAAGTGCAAAAT
CTCGTCGTTATCGAACTGGGAACAGGAAAAGCCATTCCACTGTGCGTCGATTTTCTGAACGTACGGCGAA
AAGCAAAAAAAAGGGGGGGGTTATCCCGTATTACCCCACAAGATGCAGGGCGTGCCCGAAAATGCACTTT
TTTAAGTCTAAGAAATGAAAGCGTTAGATGCACTAAAAGCGATTGA

SEQ ID NO:34 polypeptide sequence of Orf17

MKNDLNYAVELIRKADGILITAGAGMSVDSGLPDFRSVGGFWNAYPMFKEHNISFEEIATPLAYKHNQELAY
WFYGHRLVQYRNTLPHEGYQILKCWAGDKPHGYFVFTSNVDGHFQKAGFNDSHVYEVHGTLERLQCVNNCRG
LSWSASSFQPVVDNENLCLTSEKPHLPYCGGFARQNVLMFNDWSYASQYQDFKKVRLESWLKEVQNLVVIEL
GTGKAIPLCVDFLNVRRKAKKRGGLSRITPQDAGRARKCTFLSLRNESVRCTKSD.

SEQ ID NO:35 polynucleotide sequence of Orf18

TTTCTCCATAAAGAGAAATTCTTTACTTCTTACATATTTATAAAGCCTTTAATTAAGAAAAAGGAGCAAA
TAATGGCAATGAAAGTAATTATGGCAAGAGATCCACTTTTTGAGGATGTAAAAAAATATGTGCAACAACA
AAAATTTGCATCTTGCTCAATGATTCAACGCAGATTTATGTTGGGTTTTAATCGAGCTGGGCAAATTTTA
GAACAGTTGGAACAAGCGGGTATTATTTCATCAATGAAAAATGGGCAGAGAAAAGTATTATGA

SEQ ID NO:36 polypeptide sequence of Orf18

FLHKEKFFTSYIFIKPLIKKKEQIMAMKVIMARDPLFEDVKKYVQQQKFASCSMIQRRFMLGFNRAGQILEQ
LEQAGIISSMKNGQRKVL.

SEQ ID NO:37 polynucleotide sequence of Orf19

ATGTTAGTTATTAAGGAAAATAATATGAATAACCAAAACCCGATTGAAATTTACCAAACTCAAGATGGCA
CAACGCAAGTGGAAGTGAGATTTGAAAATGACACCGTTTGGCTTTCCCAAGCGCAGATGGCTATGTTATT
TGGTAAAGATATTCGCACCATCAATGAGCACATTACCAATATATTTGATGACGAAGAACTTGAGAAAGAA
TCAACTATCCGGAAATTCCGGATAGTTCGCCAAGAAGGTAAACGCCAAGTCAATCGTGAAATTGAGCATT
ATGATTTAGATATGATTATCTCTGTTGGCTATAGAGTAAAATCTAAACAAGGCATTAGTTTCCGCCGTTG
GGCAACTGCACGTTTAAAAGAATATCTGACTCAAGGCTATACCATTAACCAAAAACGTTTACAGCAAAAT
GCTCACGAATTAGAACAAGCACTTGCGCTTATTCAAAAAACGGCAAATTCATCGGAATTAACGCTAGAAA
GCGGTCGCGGATTAGTGGATATTGTCAGCCGTTATACGCATACGTTTTTATGGCTACAACAATATGATGA
AGGTTTACTTGCCGAACCACAAACACAGCAAGGCGGTACATTACCGACTTATGCTGAGGCTTTTTCTGCA
CTAGCAGAGTTAAAATCACAGCTGATGACAAAAGGTGAAGCAAGTGATCTCTTTGGACGTGAACGAGATA
ACGGCTTATCTGCGATTCTAGGTAATTTAGATCAAAGTGTATTTGGTGAACCTGCTTATCCAAGCATTGA
AGCAAAAGCGGCGCATTTACTTTATTTGTCGTCAAGAATCATCCTTTTTCAGATGGTAATAAACGTAGC
GGCGCATTTTTATTTGTAGATTTCTTACATAGAAATGGGCGTTTGTTTGATCATAATGGATACCCAGTTA

TCAATGATACTGGGCTTGCCGCGCTCACTTTATTAGTTGCTGAATCTGATCCGAAACAAAAAGAAACGCT
TATTAGGCTTATTATGCATATGCTTAAGCAAGAGAAAAAATGA

SEQ ID NO:38 polypeptide sequence of Orf19

MLVIKENNMNNQNPIEIYQTQDGTTQVEVRFENDTVWLSQAQMAMLFGKDIRTINEHITNIFDDEELEKE
STIRKFRIVRQEGKRQVNREIEHYDLDMIISVGYRVKSKQGISFRRWATARLKEYLTQGYTINQKRLQQN
AHELEQALALIQKTANSSELTLESGRGLVDIVSRYTHTFLWLQQYDEGLLAEPQTQQGGTLPTYAEAFSA
LAELKSQLMTKGEASDLFGRERDNGLSAILGNLDQSVFGEPAYPSIEAKAAHLLYFVVKNHPFSDGNKRS
GAFLFVDFLHRNGRLFDHNGYPVINDTGLAALTLLVAESDPKQKETLIRLIMHMLKQEKK.

SEQ ID NO:39 polynucleotide sequence of Orf20

ATGACAGAGAAAAATAAACCAATTTGCGTGGTATTAACGGGAGCTGGCATTAGTGCCGAAAGTGGAATTC
CAACTTTTAGATCGGAAGATGGTTTGTGGGCAGGGCATAAAGTAGAAGAAGTTTGTACGCCCGAAGCCTT
GCAAAAGAACCGTGCGAAAGTGCTTGATTTCTATAACCAACGCCGTAAAAATGCGGCAGCAGCTAAGCCA
AACGCTGCGCATCTCGCCTTAGTTGAACTAGAAAAAGCCTATGATGTGAGAATCATCACGCAAAATGTGG
ATGATTTACATGAACGTGCCGGCAGCTCGAAGGTGTTGCATTTACACGGTGAATTAAATAAAGCTCGCAG
TAGCTTTGATGAAAGTTATATTGTGGATTGTTTTGGTGATCAGAAATTAGAAGATAAAGATCCAAATGGA
CACCCAATGCGCCCTTACATCGTCTTTTTTGGTGAAATGGTGCCGATGCTAGAACGAGCGGTTGATATTG
TGGAACAAGCAGATGTTGTGTTAGTGATTGGCACTTCTTTACAAGTGTATCCAGCCAATGGCTTAGTCAA
TGAAGCCCCAAGAAAAGCGCCAATTTATCTGATTGATCCTAACCCAAATACAGGATTTGTTCGTAAGCAA
GTTATTGCAATCAAAGAAAAAGCAGGCGAGGGTGTGCCAAAAGTGGTGGCAGAGTTATTAGAGAACACCA
AAAACTCATAG

SEQ ID NO:40 polypeptide sequence of Orf20

MTEKNKPICVVLTGAGISAESGIPTFRSEDGLWAGHKVEEVCTPEALQKNRAKVLDFYNQRRKNAAAAKP
NAAHLALVELEKAYDVRIITQNVDDLHERAGSSKVLHLHGELNKARSSFDESYIVDCFGDQKLEDKDPNG
HPMRPYIVFFGEMVPMLERAVDIVEQADVVLVIGTSLQVYPANGLVNEAPRKAPIYLIDPNPNTGFVRKQ
VIAIKEKAGEGVPKVVAELLENTKNS.

SEQ ID NO:41 polynucleotide sequence of Orf21

ATGAAGAAAATTGTTTATATTGATATGGATAATGTGATGGTAGATTTTCCATCAGGTATTGCAAAACTAG
ATGATAAAACCAAGCGAGAATATGAAGGTCGATATGATGAAGTCGAGGGCATTTTTAGCTTAATGGAACC
TATGCCGAATGCGATTTCTGCGGTGCATAAATTGATGAAAAAATATCATATTTATGTGCTTTCTACTGCG
CCTTGGCATAATCCTTTTGCTTGGAGTATAAAAGTAAAATGGATTCACCATTATTTCGGTGAAGAAAAAG
GTTCAGCCTTATATAAACGATTGATTTTATCCCATCATAAAAATCTCAACCAAGGTGATTATTTAATTGA
TGATCGCACTAAAAATGGTGCTGGCAAATTTCAAGGCGAGCATGTTCATTTTGGTACAGAACAGTTTGCT
AATAAAAGGAGCCTGAAAAATGACAGAGAAAAATAA

SEQ ID NO:42 polypeptide sequence of Orf21

MKKIVYIDMDNVMVDFPSGIAKLDDKTKREYEGRYDEVEGIFSLMEPMPNAISAVHKLMKKYHIYVLSTA
PWHNPFAWSIKVKWIHHYFGEEKGSALYKRLILSHHKNLNQGDYLIDDRTKNGAGKFQGEHVHFGTEQFA
NKRSLKNDREK.

SEQ ID NO:43 polynucleotide sequence of Orf22

CATTATCGGAGTATTCACGGTAAAGAACATAAGGCACAGGTCAAGCCCTTGGCTTTGGTTCAACAAGGAC
CAAGTAGCTATTTAGTCGCACAATATGAGAATGGCGATATTTTACACCTTGCTTTGCATCGCTTGCTTAA

GGTAACAGTGAGTACAATGATATTTGAACGCCCTGATTTTAATTTGAAATCTTATGTAGAAAGCCAAAAG TTTGGTTTTACCTATGGTCGAAAAATTCGATTAACTTTCCGCATTAATAAAGATATTGGTGGATTTTTAA CAGAAACACCATTATCAATGGATCAAACAGTAAAAGATTGTGGCACTGAATATGAAATTTCCGCTACCGT GATTAAGAGCGCTATGCTGGAATGGTGGATAGCCCATTTTGGTGAAGATTACCAAGAAATTGACCGCACT TATTTAGACGAAAATGCCTAA

SEQ ID NO:44 polypeptide sequence of Orf22

HYRSIHGKEHKAQVKPLALVQQGPSSYLVAQYENGDILHLALHRLLKVTVSTMIFERPDFNLKSYVESQK FGFTYGRKIRLTFRINKDIGGFLTETPLSMDQTVKDCGTEYEISATVIKSAMLEWWIAHFGEDYQEIDRT YLDENA.

SEQ ID NO:45 polynucleotide sequence of Orf23

ATGATGAACTGGGTGCTTGGGTCAATGGAGAAAGCACCTAGCTTTCAGCATTATCATGGACATATTGATA ATATCATCAGAAGTGTTTATACGAATCCAATCTTAAGTATTGAATTGTGCAAATCTGTAACAGAAGGTAT TTGCAAAACAATTCTCAATGATAAAGGAGAAAGTATTCCTGAAAAATATCCGAATCTTGTATCTACAACA ATTAAAAAATTAGATCTGAATTATCATCAAGATTACCAATATTTGCTTGAATTAGCTAAAAGTCTGGGTT CAATTCTTCATTATGTTGCAAAAATTAGAAATGAATATGGTAGTTATGCTTCTCACGGTCAAGATATTGA ACATAAGCAAGTAAGTAGCGATCTTGCTTTATTTGTACTTCATTCAACCAATGCAATTTTAGGATTTATT CTACACTTTTACATTGCTACAAACGATTATCGAAAAAGTGAACGAATACGATATGAAGATTATGAAAGAA TCAATGAATTAATTGATGAAGAATATGAAAGGGAAGTAATATATAAAATTTCATATTCACGGGCATTATT TGATCAAGATCTAGAAGCTTATAAAGAGTTAGTACTTACATTTAAACAAACAGAACATGAGAGTCTGATG GATACGCTCTGA

SEQ ID NO:46 polypeptide sequence of Orf23

MMNWVLGSMEKAPSFQHYHGHIDNIIRSVYTNPILSIELCKSVTEGICKTILNDKGESIPEKYPNLVSTT IKKLDLNYHQDYQYLLELAKSLGSILHYVAKIRNEYGSYASHGQDIEHKQVSSDLALFVLHSTNAILGFI LHFYIATNDYRKSERIRYEDYERINELIDEEYEREVIYKISYSRALFDQDLEAYKELVLTFKQTEHESLM DTL.

SEQ ID NO:47 polynucleotide sequence of Orf24

ATGAATGATTGGAAGGTTATAACTTTAGCTGATTGCGCTTCATTTCAAGAAGGTTATGTTAATCCATCAA AAAATGAACCAAGCTACTTTGGAGGAACAATTAAATGGTTGAGAGCAACAGATTTAAACAATGGTTTTGT ATATAAAACCTCTCAAACTTTAACAGAAAAAGGATTTTTAAGTGCAAAGAAGAGTGCTGTATTATTTGAA CCAGATAGTTTAGCAATTAGCAAATCAGGAACTATTGGACGAATTGGAATCTTAAAAGATTACATGTGTG GAAATAGAGCTGTAATTAATATCAAAGTTAATGAAAATATTTGTAACCCATTATTTATTTTTTATACCTT ATTAAATAGCAAAGAACAAATTGAAACTTTAGCTGAAGGTAGTGTCCAAAAAAATCTATATGTATCAGCT TTAAGTAAAGTTAAATTATTACTTCTAGATATAAATAAGCAAAAGGAAATTGGATATATTCTAAATACTT TAGATCAAAAAATAGAACTCAACACTCAAATCAACCAAACCTTAGAACAAATCGCCCAAGCCCTGTTTAA AAGCTGGTTTGTCGATTTCGATCCCGTGCGTGCCAAAATCCAAGCCCTTTCAGACGGTCTTAGCCTTGAA CAAGCAGAACTTGCCGCCATGCAGGCAATCAGCGGAAAAACACCCGAAGAACTGACCGCACTTTCACAAA CACAGCCTGACCGCTACGCCGAACTAGCCGAAACCGCCAAAGCGTTTCCGTGTGAGATGGTGGAGGTTGA TGGGGTTGAAGTGCCGAAGGGGTGGGAATTATCTACGATTGGCGATTGTTATGATGTCGTTATGGGGCAA TCTCCAAAAGGAGAAACTTATAATGAAAACAAACAAGGGATGCTTTTCTATCAAGGTCGTGCAGAATTTG GTTGGCGCTTTCCTACCCCAAGATTATTTACAACAGATCCTAAACGTATTGCAGAACAAAATTCTATTTT

AATGAGCGTTCGAGCTCCTGTTGGGGACATTAATATAGCACTTGAAAAATGCTGTATTGGTCGCGGATTA
GCTGCATTACAACATAAGAGTAAAAGTTTGTCGTTCGGTTTATATCAAATACAATCTATAAAACCAGAAT
TAGATTTATTTAATGGTGAAGGAACTGTTTTTGGTTCTATCAATCAGGATAACTTAAAAAATATCCAAAT
TATTAACCCTGATGAAAAATTTATTCAGCTTTTTGAAAAATATTTATCATCTTGTGATTCAAAAATTATG
AATAACGAGATAGAAAATAATGCACTGAAAGAAATAAGGGATTTATTGTTACCTAGATTATTGAGTGGAG
AAATTCAATTATGA

SEQ ID NO:48 polypeptide sequence of Orf24

MNDWKVITLADCASFQEGYVNPSKNEPSYFGGTIKWLRATDLNNGFVYKTSQTLTEKGFLSAKKSAVLFEPD
SLAISKSGTIGRIGILKDYMCGNRAVINIKVNENICNPLFIFYTLLNSKEQIETLAEGSVQKNLYVSALSKV
KLLLLDINKQKEIGYILNTLDQKIELNTQINQTLEQIAQALFKSWFVDFDPVRAKIQALSDGLSLEQAELAA
MQAISGKTPEELTALSQTQPDRYAELAETAKAFPCEMVEVDGVEVPKGWELSTIGDCYDVVMGQSPKGETYN
ENKQGMLFYQGRAEFGWRFPTPRLFTTDPKRIAEQNSILMSVRAPVGDINIALEKCCIGRGLAALQHKSKSL
SFGLYQIQSIKPELDLFNGEGTVFGSINQDNLKNIQIINPDEKFIQLFEKYLSSCDSKIMNNEIENNALKEI
RDLLLPRLLSGEIQL.

SEQ ID NO:49 polynucleotide sequence of Orf25

ATGGAATTAATAAGCGATAATCCAATAAAAGATTCTAGCAATGATTTATTAGGTAGAGCTAGTAGTGCAG
AAGCATTTGCTAAACACATTTTTTCATTTGACTATAAAGAAGGTTTGGTTGTGGGATTATGTGGAGAATG
GGGAAATGGTAAAACATCCTATATAAATTTAATGCGACCAGAATTAGAAAAAAATTCTTTTGTACTTGAT
TTTAATCCTTGGATGTTTAGTGATGCTCATAACTTAGTTGCTTTATTTTTTACTGAAATCTCTGCTCAGT
TAAGAGATTATGAGGATGATAATGAGCTAATTGATAGTTTGAGTAGTTTTGGAGAGTTGTTATCTAATTT
AAAACCTATTCCATTTGTAGGAAATTATTTAGTGTCTTGGGTGGCTGTTTAAGTTTTTTTCAAAGAAA
AAGAAAGAAAAAAACAGTTTGAAAAATCAACGTGATAAATTAATTAAAGTTCTAAAGGAAATAAGTAAAC
CTATTACTGTAATTTTAGATGATATAGACCGTTTATCATCTGATGAATTACAATCAATTCTAAAATTGGT
CAGAGTTACAGGAAACTTTCCTAATATTGTTTATGTTTTATCATTTGATAAAAAATAGAGTAATTAAACCA
TTAAATGATAATACCATTGATGGCCAGGATTATTTAGAGAAGATAATTCAGATTCCATTCGATATACCAC
AGGTACCTAAAAAACTATTACAAGAAAATTTATTTCATCTTTAGATAAGATTTTAAGGGATGTTTACCT
AGATAAGGCGCGTTGGTCTAATGCATATTGGAATATCATTAAGCCAACAATAAAAAATATTCGAGATATT
AAGCGTTACACATCTTCTCTATCGAATATCTTTAAACAATTAGGTAAAGAAATTGATGTGGTTGATTTAC
TCACTATTGAAGCGATAAGAATTTTCTTTCCAGATAAATTTAAAGAAATTTTTGAACTTAAAGATTATCT
CTTGGCACGATCAGATAATGACAAAAGAAAAGTTAAGTTAAGTGATTTTATTCAAGATAATGAAATGTAT
GAGTCTTTTCTAGAAGTTTTATTTGATATTGATAATATAAATTCAAATAATGAATTCCTAAAAAAATAGAA
GGATTGCTTATTCGGCATTCTTTGATTTATATTTGAACAAGTTATGAGTCCTGAGTTCATAAATGTTAA
ATTATCACAAAAAGTTTGGCTTGCAATGCAGTCAGAAGAAGATTTCAAGATCGCTTTATCAGCTGTTCCT
GACGATTCTCTAGAAAATGTAGTTAACAATTTAATTGACTATGAAAAAGACTTTACTAAAGAAATAGCTC
TAGCAACTATACCAACATTATATAGAAATTTACCAAGAGTGCCTGAAAAAGAATTAGGATTCTTTGACTT
TGGGGCGGATATGGTTTGGAGTCGCTTAGTTTATAGATTACTTAGAAGACTTCCTGAGAAGGATAAAAAA
GAAGTTATTACTCAACTATTAAATTCTAGCGATCTATATGGGCAATATCAAATTGTAGGAATTATTGGAT
ATCGAGAGGGCCGAGGTCATCAATTAGTATCTGAATCGGATGCAAAAGACTTGGAGGAAATATTTTTAAA
TAATATTCGCTCTGCAACAATTAAAGAACTTGCAGGAACCTATAATTTGTCACATATAATCTATTTCTTT
GTTTCAATTGGAAACCCTTTTTCTGATGATATATTAAGTTCCCCTGAAGTATTTTTATCATTACTTAAAT
CTTCAATATCAGAACGTAAATCTCAAAGAGGGGATGATCCTACAATACATAGAGAGAAAATTCTACTTTG
GGATGCCTTAATTAAAATTTGTGGAGATGAGGATAAAGTAAATAGTTTAATTGAAAAAATAGCTGAAGAT
GAAGAACTTAGAAATAAAGATTATATGGAACTTGCAATTAAATATAAGAATGGATACCGACATAAAAAAT
CAATGAATCATGAAGATGATTTAGATGAGTTTTAA

SEQ ID NO:50 polypeptide sequence of Orf25

MELISDNPIKDSSNDLLGRASSAEAFAKHIFSFDYKEGLVVGLCGEWGNGKTSYINLMRPELEKNSFVLDFN
PWMFSDAHNLVALFFTEISAQLRDYEDDNELIDSLSSFGELLSNLKPIPFVGNYFSVLGGCLSFFSKKKKEK
NSLKNQRDKLIKVLKEISKPITVILDDIDRLSSDELQSILKLVRVTGNFPNIVYVLSFDKNRVIKPLNDNTI
DGQDYLEKIIQIPFDIPQVPKKLLQENLFSSLDKILRDVYLDKARWSNAYWNIIKPTIKNIRDIKRYTSSLS
NIFKQLGKEIDVVDLLTIEAIRIFFPDKFKEIFELKDYLLARSDNDKRKVKLSDFIQDNEMYESFLEVLFDI
DNINSNNEFLKNRRIAYSAFFDLYFEQVMSPEFINVKLSQKVWLAMQSEEDFKIALSAVPDDSLENVVNNLI
DYEKDFTKEIALATIPTLYRNLPRVPEKELGFFDFGADMVWSRLVYRLLRRLPEKDKKEVITQLLNSSDLYG
QYQIVGIIGYREGRGHQLVSESDAKDLEEIFLNNIRSATIKELAGTYNLSHIIYFFVSIGNPFSDDILSSPE
VFLSLLKSSISERKSQRGDDPTIHREKILLWDALIKICGDEDKVNSLIEKIAEDEELRNKDYMELAIKYKNG
YRHKKSMNHEDDLDEF.

SEQ ID NO:51 polynucleotide sequence of Orf26

TATGACAAAAGTTTAGACAAAATTGCAAAACAATTAAGAGATTCTGATAAAAAGGTTAATCTAATTTACG
CCTTTAATGGAAGTGGAAAAACCCGTTTATCAAAAGTCTTTAAGAATCTTATTGCACCTAAAGAAAATCA
TGACAATGAAGAAGATCTAACACGAAGAAAAATTCTTTATTTCAATGCCTTTACCGAAGATTTATTCTAT
TGGGATAATGATCTACTTAATGACACAGAACCAAAATTAAAGATTCAACCAAATTCTTTTATTCGCTGGT
TGATTAGAGATCAAGGGGATGAAGGTAAAGTAATTGGAAAATTTCATCATTATTGTGATGAAAAACTTAT
GCCTAAATTTGATATAGAAAATAATCAAATTACATTCAGTTTTGCACGTGGAGATGATACGCCTGAAGAA
AATATAAAACTATCGAAGGGGGAAGAAAGTAATTTTATTTGGAGTATTTTTCATACGTTAATTGAACAAG
TTGTTGCAGAATTAAATATCTCAGAGCCTAGTGAACGCACTACTAATGAATTTGATGAACTTAAATATAT
CTTTATTGATGATCCAGTAAGTTCATTGGATGAAAATCATCTTATTCAATTAGCTGTTGATTTAGCAGAA
TTAGTCAAAGATAGTCCCGATACTATAAAATTTATTATCACCACACACAATCCTTTATTTTATAACGTTT
TATACAATGAACTTGGAGCAAAAAATGGTTATATTCTAAGAAAAGATGAAAATAAGAATGAAAAAGAAAG
ATTTGATCTTGAGGTGAAACAAGGTGGTTCAAACAAGAGTTTCTCCTATCATCTTTTTCTAAAAAATCTA
CTTGAAGAAGTTGAACCTAAAGATATTCAAAAATATCACTTCATGTTACTGAGAAATTTATATGAAAAAG
CTGCTAACTTTCTTGGTTATTCAGGATGGTCAAATCTATTACCCAATGATGATGCAAGACAAAGCTATTA
CACTCGTATAATCAATTTTACTAGTCACTCTACGTTATCAAATGAGATAATCGCTGAGCCAACAGATGCC
GAAAAGAAGATTGTTAAATATTTACTTGAACATCTAATTAATAATTATGGTTTCTATATAGAAGAAAATA
TTAAAGACCCACAAACTGATAATATAACAGAGTAA

SEQ ID NO:52 polypeptide sequence of Orf26

YDKSLDKIAKQLRDSDKKVNLIYAFNGSGKTRLSKVFKNLIAPKENHDNEEDLTRRKILYFNAFTEDLFY
WDNDLLNDTEPKLKIQPNSFIRWLIRDQGDEGKVIGKFHHYCDEKLMPKFDIENNQITFSFARGDDTPEE
NIKLSKGEESNFIWSIFHTLIEQVVAELNISEPSERTTNEFDELKYIFIDDPVSSLDENHLIQLAVDLAE
LVKDSPDTIKFIITTHNPLFYNVLYNELGAKNGYILRKDENKNEKERFDLEVKQGGSNKSFSYHLFLKNL
LEEVEPKDIQKYHFMLLRNLYEKAANFLGYSGWSNLLPNDDARQSYYTRIINFTSHSTLSNEIIAEPTDA
EKKIVKYLLEHLINNYGFYIEENIKDPQTDNITE

SEQ ID NO:53 polynucleotide sequence of Orf27

ATGAACGACTTAATCATCTACAACACTGACGATGGTAAATCTCACGTTGCTTTATTAGTTATCGAAAATG
AGGCTTGGCTGACTCAAAATCAGCTTGCGGAACTTTTTGACACCTCTGTACCAAATATAACCACTCATAT
AAAAAACATATTACAAGACAAAGAGTTAGATGAGTTTTCAGTTATTAAGGATTACTTAATAACTGCCCAA
GATAGCAAACAATATCAAGTAAAACATTATTCCCTTGATATGATTCTCGCCATCGGCTTTCGTGTGCGCA
GCCCTCGTGGTGTACAGTTTCGTCGTTGGGCGAATACGCAATTACGTACTTATTTAGATAAAGGTTTTCT
ATTAGATAAAGAGCGGTTGAAAAATCCTCAAGGTCGATTTGATCATTTTGATGAATTACTGGAACAAATT

CGCGAAATTCGAGCCAGTGAATTGCGGTTTTATCAAAAAGTACGAGAGTTATTTAAATTATCCAGTGACT
ACGATAAAACAGATAAAGTCACTCAAATGTTTTTGCAGAAACACAAAATAAGTTGATTTATGCCATTAC
ACAACAAACCGCCGCAGAGCTTATTTGTACGCGTGCAAATGCCAAATTGCCTAATATGGGTCTTACCTCT
TGGAAAGGTGCTGTTGTACGTAAAGGCGATATTATTACCGCTAAAAACTATTTAACTCATGATGAATTAG
ATTCTTTGAATCGTTTAGTGATGATCTTTTTAGAAAGTGCTGAATTACGCGTTAAAAATCGTCAAGATCT
CACATTAAATTTCTGGCGTAATAATGTCGATAATTTAATTGAATTTAACGGTTTTCCGTTGCTTATCGGT
AATGGAACCCGAACCGTAAAACAAATGGAAACCTTTACCAAAGAACAATATGCCTTATTTGATCAGGTCA
GAAAACAACAAAAACGCATACAAGCTGATAATGAAGATTTAGAAATTTTAGAAAACTGGCAGAAAGATCT
GAAAAAGCAAAAGCATTAA

SEQ ID NO:54 polypeptide sequence of Orf27

MNDLIIYNTDDGKSHVALLVIENEAWLTQNQLAELFDTSVPNITTHIKNILQDKELDEFSVIKDYLITAQ
DSKQYQVKHYSLDMILAIGFRVRSPRGVQFRRWANTQLRTYLDKGFLLDKERLKNPQGRFDHFDELLEQI
REIRASELRFYQKVRELFKLSSDYDKTDKVTQMFFAETQNKLIYAITQQTAAELICTRANAKLPNMGLTS
WKGAVVRKGDIITAKNYLTHDELDSLNRLVMIFLESAELRVKNRQDLTLNFWRNNVDNLIEFNGFPLLIG
NGTRTVKQMETFTKEQYALFDQVRKQQKRIQADNEDLEILENWQKDLKKQKH

SEQ ID NO:55 polynucleotide sequence of Orf28

ATGCAACAGCGTGTACTTTTTTTAAAAGCGTGGCTAAGCCAACGTTATACTAAAACTGAACTGTGTCAGC
AGTTTAATATTAGCCGTCCAACGGCAGATAAATGGATTAAACGCCACGAACAGCTTGGTTTTGAGGGCTT
AAGCGAGTTATCTCGTAAATCTTATCATAGCCCTAATGCCACGCCACAATGGATTTGTGACTGGCTTATC
AGTGAGAAACTTAAACGTCCTCACTGGGGTGCCAAAAAGCTTTTAGATAACTTTACTCGGCATTTTCCAG
AAGCGAAAAAGCCGTCTGATAGCACGGGCGATTTAATTTTGGCGTGTGCAGGGTTAAAACGTCGTATGAG
TGCAGACACACAATCTTTTGGCGAATGCATCGCACCCAATACCACCTGGAGTGCTGACTTCAAGGGGCAA
TTTTTACTCGGCAATCAGAAGTTCTGCTATCCGCTGACGATTACAGATAATTTCAGTCGCTTTTTATTTT
GTTGTAAGGGGTTGCCGAATACAAAATCAGCGCCTGTTATTGCTGAGTTTGAACGTCTTTTTGAGCAATT
TGGTCTGCCGTATTCGATTCGTACCGATAACGATTCATCTTTTGCATCACAAGCATTAGGTGGATCTAGG
TGTATTGACTTAGGTATTCCTTCTGAACGAATTAAGCCATCACACCCAGAGCAGAACGGACGACACGAGC
GAATGCACCGTAGCTTAAAAACAGCGCTTCAACCTCAAAATAGCTTTGAAGCTCAACAGACATTCTTCAA
CCAATTCTTACGAGAATACAAAGAAGAATGTTCACACGAAGGCGTTTGA

SEQ ID NO:56 polypeptide sequence of Orf28

MQQRVLFLKAWLSQRYTKTELCQQFNISRPTADKWIKRHEQLGFEGLSELSRKSYHSPNATPQWICDWLI
SEKLKRPHWGAKKLLDNFTRHFPEAKKPSDSTGDLILACAGLKRRMSADTQSFGECIAPNTTWSADFKGQ
FLLGNQKFCYPLTITDNFSRFLFCCKGLPNTKSAPVIAEFERLFEQFGLPYSIRTDNDSSFASQALGGSR
CIDLGIPSERIKPSHPEQNGRHERMHRSLKTALQPQNSFEAQQTFFNQFLREYKEECSHEGV.

SEQ ID NO:57 polynucleotide sequence of Orf29

TGCCAAACGGCGAACAAATCCGCAGAATTAAGCAGCGTTGTGGCTATTCTCGCTTCATGTTTAATCGGGT
TAACTTGGCAGAATGAACAATATAAGCAAGATAATGGCGTCAAGTTCAGTTATACGAAAATCGCCAAATT
GCACCACAAAGTCACCAATACCCACAAAAAAAACTACTTGCATCAAATCCCACACCGAATCAGCAAAAAC
CACGCAATGATTTATATTGAGAGTTTGCAAGCAACAAATTACCAAGGAGATGCGGAAAATACAGTAAAAC
GCGAAACAAAAATCAGACTTAAACCGTTCAACTTCAGCACAATCTTGGCATGA

SEQ ID NO:58 polypeptide sequence of Orf29

CQTANKSAELSSVVAILASCLIGLTWQNEQYKQDNGVKFSYTKIAKLHHKVTNTHKKNYLHQIPHRISKN
HAMIYIESLQATNYQGDAENTVKRETKIRLKPFNFSTILA

SEQ ID NO:59 polynucleotide sequence of Orf30

TTGCAATTAAAAAAATTTATTTTAGAAACTCCTGAAAATATTCTAACTGAACTTTGGGGAAATTACATTA
AAGATGATCGTATAACTCAATGGGCAAATTTAGTGTTATCTTATTGTAAACCTTCAAACCACAATGAAAT
GAAATTAATTTTGACAAAAATTGTAAATGAAAAAACAATTTTTAATGATAAAGATGATGTAAACAAATTA
GAAGAAATGGCAAAAATATACATAACCAATCAGAAAATTAATAGTTTATAA

SEQ ID NO:60 polypeptide sequence of Orf30

LQLKKFILETPENILTELWGNYIKDDRITQWANLVLSYCKPSNHNEMKLILTKIVNEKTIFNDKDDVNKL
EEMAKIYITNQKINSL

SEQ ID NO:61 polynucleotide sequence of Orf31

ATGATTTTCTCTAAAAATAAGTATCCACCTTTACATGAATTCACGTCATTAATGAATAGAGTCGATAATT
TTCTTAATCATGATGCAGAAAATAGGGTTGCATACTATAAGAAACGTAGTGGTATTGATTTAGAAAAAGA
TGTATATGAGGCTATTTGTTATTGTGCTCAAAATACTCCTTTCGAAGACACTATTAGTTTAGTATCAGGG
AAACATTTTCCAGACATTGTAGCTAGTCAATATTATGGTATTGAAGTAAAAAGTACACAAGGAGATAAAT
GGACTTCAATTGGCAGTTCTATTCTTGAGTCTACACGAATTCCAAATATAGAAAAAATTTTCTTAACATT
TGGTAAATTAGGTGGAAATATTAAATTCCTATCCAAACCATATGAGTCGTGTTTATGTGATATAGCTGTA
ACCCATTACCCTAGATATAAAATAGATATGTTATTAGAAAAGGGGGAGAGCATATTTGAAAAAATGGAGA
CCACATATGATTCTCTCCGAGAATTAGATAATCCAATAACTCCTGTAGCTAAATACTATAAATCTCTATT
AATAGAAGGTGAAAGTTTATGGTGGACTTCAAACAATGTTTTAGATGATATTGCCCCTCCCAAAGTTAGA
CACTGGAAGGTAATAGAAAAATATGAGCGAGATATGTTAATTGCTCAAGCATATGCTTTCTTCCCTGAAA
CGATCTTAGGAAATCCTAGAAATAAATATGATAAATTCGCACTATGGCTAGTGACTAAACATGGAGTAAT
AAACACTAGTTTAAGAGATGAGTTTTCTGCAGGAGGGCAACAAAAAATAACTGATACTTGTGGTGAAACA
CATCTTTGTTCTGCTGTATTAAAGAGAGTAGAGAACAATATTCTTGCAATTAAAAAAATTTATTTTAGAA
ACTCCTGA

SEQ ID NO:62 polypeptide sequence of Orf31

MIFSKNKYPPLHEFTSLMNRVDNFLNHDAENRVAYYKKRSGIDLEKDVYEAICYCAQNTPFEDTISLVSG
KHFPDIVASQYYGIEVKSTQGDKWTSIGSSILESTRIPNIEKIFLTFGKLGGNIKFLSKPYESCLCDIAV
THYPRYKIDMLLEKGESIFEKMETTYDSLRELDNPITPVAKYYKSLLIEGESLWWTSNNVLDDIAPPKVR
HWKVIEKYERDMLIAQAYAFFPETILGNPRNKYDKFALWLVTKHGVINTSLRDEFSAGGQQKITDTCGET
HLCSAVLKRVENNILAIKKIYFRNS

SEQ ID NO:63 polynucleotide sequence of Orf32

CTGTTGGGCCCCAACAATTCCGATTCTGAACATCATGGTAATATTGAAAATCGTAGGCTAAGCATAGAGCAT
GAAGGGAAATATATTAACGAATTATCTAAAGGCATGCTCGAACGTCGTCTTACTATAAGAGAATGTGCTAGA
TTACAAACGTTTCCTGATAGATACCAATTTATTTTACCTAAAACAGCAGAAAACGTTTCTGTTTCAGCCAGT
AATGCCTATAAAATTATTGGCAATGCGGTACCATGTATATTAGCTTATAATATTGCTAAAAATATAGAAAAA
AAATGGAATCTTTATTTTAAATAG

SEQ ID NO:64 polypeptide sequence of Orf32

FLLGPNNSDSEHHGNIENRRLSIEHEGKYINELSKGMLERRLTIRECARLQTFPDRYQFILPKTAENVSV
SASNAYKIIGNAVPCILAYNIAKNIEKKWNLYFK

SEQ ID NO:65 polynucleotide sequence of Orf33

ATGAGTGTACTCAGTTACGCACAAAAAAATCGGTCAAGCCTTAATGGTGCCTGTGGCAGCCTTACCTGCTG
CTGCATTATTAATGGGTATTGGCTATTGGATCGACCCAGATGGTTGGGGTGCAAATAGTCAATTAGCCGC
ATTATTAATTAAATCTGGCGCAGCAATTATTGACAACATGGGCTTACTCTTCGCTGTGGGCGTCGCTTTT
GGGCTTGCAAAAGATAAACACGGTTCCGCCGCACTTTCAGGCCTTGTTGGTTTCTACGTAGTAACCACCC

TACTTTCCCCTGCTGGTGTAGCACAATTACAACACATTGATATTAGTGAAGTGCCTGCCGCATTCAAAAA
AATCAATAACCAATTTATTGGGATTTTAATTGGTGTGATTTCAGCTGAACTTTACAACCGTTTCTATCAA
GTTGAATTACCAAAGGCACTTTCGTTCTTTAGCGGAAAACGCCTCGTCCCAATTTTGGTTTCTTTCGTGA
TGATCGCCGTATCATTTGCCTTACTCTATATTTGGCCTCATATTTTTAACGCTCTCGTTTCATTTGGTGA
ATCCATCAAAGATTTAGGTGCAGTAGGTGCGGGGATCTACGGTTTCTTCAACCGCTTATTAATTCCTGTA
GGCTTACACCATGCCTTAAACTCTGTATTCTGGTTTGATGTAGCGGGTATCAACGATATTCCAAACTTCT
TGGGCGGCGCTAAATCCATTGCCGAAGGCACTGCAACCGTGGGGCTAACTGGTATGTATCAAGCTGGTTT
CTTCCCTGTCATGATGTTTGGTTTACCAGGTGCTGCTCTTGCAATTTATCACTGCGCAAAACCAAACCAA
AAAGTACAAGTGGCCTCAATTATGCTTGCGGGTGCGTTAGCCTCTTTCTTTACAGGGATCACTGAACCGC
TTGAATTCTCATTTATGTTCGTTGCACCTGTACTTTATGTATTGCATGCATTATTAACAGGTATCTCTGT
ATTCATTGCAGCTACAATGCACTGGATTGCAGGATTCGGATTTAGTGCAGGTTTAGTGGATATGGTACTT
TCTAGCCGTAACCCACTTGCCGTTAGCTGGTATATGTTACTTGTACAAGGTATTGTATTCTTTGCTATCT
ATTATTTTGTGTTCCGTTTTGCAATTAATGCCTTTAATCTCAAAACGCTAGGACGTGAAGATAAAGCGGA
AACAGCTGCAGCCCCAACTCAAAGCGACCAATCTCGCGAAGAAAGAGCGGTGAAATTTATTGCTGCTTTA
GGTGGTTCAGAAAACTTCAAAACTGTGGATGCTTGTATCACTCGTTTACGCTTAACTTTAGTTGATCATC
ACAATATTAACGAAGATCAACTTAAAGCGCTTGGTTCAAAAGGTAATGTAAAATTAGGCAATGATGGATT
ACAAGTCATTTTAGGGCCTGAAGCTGAACTTGTGGCAGATGCGATTAAAGCAGAATTAAAATAA

SEQ ID NO:66 polypeptide sequence of Orf33

MSVLSYAQKIGQALMVPVAALPAAALLMGIGYWIDPDGWGANSQLAALLIKSGAAIIDNMGLLFAVGVAF
GLAKDKHGSAALSGLVGFYVVTTLLSPAGVAQLQHIDISEVPAAFKKINNQFIGILIGVISAELYNRFYQ
VELPKALSFFSGKRLVPILVSFVMIAVSFALLYIWPHIFNALVSFGESIKDLGAVGAGIYGFFNRLLIPV
GLHHALNSVFWFDVAGINDIPNFLGGAKSIAEGTATVGLTGMYQAGFFPVMMFGLPGAALAIYHCAKPNQ
KVQVASIMLAGALASFFTGITEPLEFSFMFVAPVLYVLHALLTGISVFIAATMHWIAGFGFSAGLVDMVL
SSRNPLAVSWYMLLVQGIVFFAIYYFVFRFAINAFNLKTLGREDKAETAAAPTQSDQSREERAVKFIAAL
GGSENFKTVDACITRLRLTLVDHHNINEDQLKALGSKGNVKLGNDGLQVILGPEAELVADAIKAELK

SEQ ID NO:67 polynucleotide sequence of Orf34

ATGAAAACAACTTCTGAAGAATTAACGGTATTTGTGCAAGTAGTCGAAAATGGCAGTTTCAGCCGTGCAG
CCAAGCAGCTATCAATGGCAAATTCTGCGGTAAGTCGTGTGGTGAAAAGGCTAGAAGAAAAATTGGGTGT
GAACCTAATCAACCGCACTACTAGACAGCTTAGACTAACAGAAGAAGGCTTACAATATTTTCGTCGCGTA
CAGAAAATTCTGCAAGATATGGCTGCAGCTGAAGCTGAAATGTTGGCAGTGCACGAAGTCCCACAAGGCA
TACTACGCGTAGATTCAGCCATGCCGATGGTGTTACATCTGCTAGTGCCACTGGCAGCAAAATTCAACGA
ACGCTATCCGCATATCCAACTTTCGTTAGTTTCTTCTGAAGGCTATATCAATCTGATAGAACGCAAAGTC
GATATTGCCTTACGAGCTGGAGAATTGGATGATTCTGGGCTGCGTGCTCGTCATCTATTTGATAGCCACT
TCCGCGTAATCGCCAGTCCAGACTACTTGGCAAAACACGGCACGCCACAATCAACTGAAGCTCTTGCCAA
CCATCAATGTTTAGGCTTCACTGAGCCCAGTTCACTAAATACATGGGAAGTTTTAGATGCTCAAGGAAAT
CCCTATAAAATCTCACCGTACTTTACCGCCAGCAGCGGTGAAATTTTACGGTCATTGTGTCTTTCAGGCT
GTGGTATTGCTTGCTTATCAGATTTTTTGGTAGACAATGACATCGCTGAAGGAAAATTAATTCCCTTACT
TACTGAACAAACCGCCAATAAAACGCTCCCCTTCAATGCTGTTTACTACAGCGATAAAGCAGTCAACCTT
CGCCTACGTGTGTTTTTAGACTTTTTAGTAGAAGAGCTAAGGGGATAA

SEQ ID NO:68 polypeptide sequence of Orf34

MKTTSEELTVFVQVVENGSFSRAAKQLSMANSAVSRVVKRLEEKLGVNLINRTTRQLRLTEEGLQYFRRV
QKILQDMAAAEAEMLAVHEVPQGILRVDSAMPMVLHLLVPLAAKFNERYPHIQLSLVSSEGYINLIERKV

DIALRAGELDDSGLRARHLFDSHFRVIASPDYLAKHGTPQSTEALANHQCLGFTEPSSLNTWEVLDAQGN
PYKISPYFTASSGEILRSLCLSGCGIACLSDFLVDNDIAEGKLIPLLTEQTANKTLPFNAVYYSDKAVNL
RLRVFLDFLVEELRG

SEQ ID NO:69 polynucleotide sequence of Orf35

AGAGCATTAGTAGAGAATAAAAAGGAGTTCGAAAATTTAAAAAACTCACTGATTACACTCAAAAAATCTT
ATAACGACGCACAAGAACAAATAACTGAAATTTCCCAGTGGCACGAACAGTCAGAGAAATTAAGTGGCGA
CATTTCGAACTATGAATTCACCGCACAAAATAATCTTACTAAAATTACGACATTAGCAACCACAGCGGGA
AAACCAATAAACCCCAAATCGgAAAAATATCATGAAGATATTGAAGGTATGATTAAATTATTCAATAAAC
AAAAAGAGGAGATTGAAATGATTATTGAAGACGCCAACCGAGCAAGCATGGCAGGTTCGTTTAAAACTCA
ATCTGAAAATATCGATAGTAAAATGAAAGCTGTAGATAAAATTTTGcCTTgGGGTCACTTgGTTGCAACA
TCTGTTATTTCATTGTTCAATTATTCAACAAGCCTGAGTGCAGCAGACAGCCTTAATATTTTACAATTTC
TTGCTAAGTCCATTGTGACAATCCCGTTACTTGTCATCGCCTGGTTGAAAGCAAAAGAACGGGCTTATCT
CTTTAGATTAAGGGAGGATTATAACTACAAATATTCCTCAGCAATGGCATTTGAAGGTTATAAGAAACAA
GTACAAGAACAAGACCCTAAATTACATCAGCAACTTCTGCAAATTGCCGTGGATAATTTGGGGATAAATC
CAACCAAAGTCTTTGACAAAGATTTAAAAAGCACACCACTTGAAACAATTATCGATGGAGTAGGAAAACG
CCTGGATAAAGCTGTTGATGGTATTAAAGGAGAGGTGAATGACATTCCAAAGAAAAcCAAAAGAATTAAT
TGA

SEQ ID NO:70 polypeptide sequence of Orf35

RALVENKKEFENLKNSLITLKKSYNDAQEQITEISQWHEQSEKLSGDISNYEFTAQNNLTKITTLATTAG
KPINPKSEKYHEDIEGMIKLFNKQKEEIEMIIEDANRASMAGSFKTQSENIDSKMKAVDKILPWGHLVAT
SVISLFNYSTSLSAADSLNILQFLAKSIVTIPLLVIAWLKAKERAYLFRLREDYNYKYSSAMAFEGYKKQ
VQEQDPKLHQQLLQIAVDNLGINPTKVFDKDLKSTPLETIIDGVGKRLDKAVDGIKGEVNDIPKKTKRIN

SEQ ID NO:71 polynucleotide sequence of Orf36

GATTATATGTTATCAGCAACGCAATTTCTTGTTTTAGAAAAAGCACTTAGTAAGGAAAGATTATCTACAT
ACAAAAACTATGTGAAAAATAAAACTTCAGAAAGTATTAATGATAACATGGTTGCTTTATATGAATGGAA
TTCTGAAATAGCGGGCTATTTTCTTGAATTCTGTAATATATATGAGATTTCATTAAGAAATGCTATTTAT
AGATCAATAGATTCGTATGATCATTATGGTATCAGACAGAGACAAATACTTAGACAAAGTCCTAAATTAA
GAGAAAAAGTTGAAGAATTAGGTAGAAATGCGACTGATGGAAAAATCATATCTAGTTTACATTTTCACTT
TTGGGAATTTTTTGAAGAAGTTTTTCTTGTGGAATTCTCGTGA

SEQ ID NO:72 polypeptide sequence of Orf36

DYMLSATQFLVLEKALSKERLSTYKNYVKNKTSESINDNMVALYEWNSEIAGYFLEFCNIYEISLRNAIY
RSIDSYDHYGIRQRQILRQSPKLREKVEELGRNATDGKIISSLHFHFWEFFEEVFLVEFS

SEQ ID NO:73 polynucleotide sequence of Orf37

ATGAAACTAATATCTCTATTCTCAGGTTGTGGGGGAATGGATATCGGATTTGAAGGTAATTTCTCTTGTC
TAAAAAAATCTATTAATGAGGAGCTCCACCCTGAATGGATCAGCTCCACAGAAAATGAATGGGTTACCGT
TTCGCCCACCTCTTTTGAGACAATTTTTGCTAATGATATTAAACCTGATGCTAAAGCAGCATGGGTTTCT
TATTTCTTAGACCAAAAAGCGAATGCAAACGAAATCTACCACTTAGAAAGCATTGTTGATCTTGTAAAAA
AAGAACGGGAAACTCACAATATTTTCCCAAAAGGCATTGATATATTAACAGGTGGATTTCCTTGTCAAGA
TTTTTCTGTAGCCGGAAAACGATTAGGATTTGATTCTCACAAAAATCATCATGGAAAAAATATCAAATATA

GATGAACCCTCAATTGAAAATAGAGGACAATTATACATGTGGATGAGAGAAGTAATATCTATAACTCACC
CCAAATTATTCATAGCTGAAAATGTAAAAGGATTAACGAACCTTAAAGATGTAAAAGAAATTATTGAACA
TGATTTTGGTCAAGCTAGTGACGAAGGATACTTAATTGTACCAGCTTCAGTATTAAATGCTCAGTTTTAT
GGAGCTCCTCAATCACGTGAGCGTGTCATTTTTTTTTGGTTTTAA

SEQ ID NO:74 polypeptide sequence of Orf37

MKLISLFSGCGGMDIGFEGNFSCLKKSINEELHPEWISSTENEWVTVSPTSFETIFANDIKPDAKAAWVS
YFLDQKANANEIYHLESIVDLVKKERETHNIFPKGIDILTGGFPCQDFSVAGKRLGFDSHKNHHGKISNI
DEPSIENRGQLYMWMREVISITHPKLFIAENVKGLTNLKDVKEIIEHDFGQASDEGYLIVPASVLNAQFY
GAPQSRERVIFFWF

SEQ ID NO:75 polynucleotide sequence comprising orfs1, 2, 3, 4, 5, 6, 7, 8 and non-coding flanking regions of these polynucleotide sequences.

TATTGCAAACACTTCTCAGATGATTAAATAACATGGATACACGTTTGCCCACACGGATTGCTGGTAACCTTT
GACAGTCGATGAAATAGGTGTGCTATGAGCCATTTATTTATTACCCAATGATGTGCAATGAAAAGATAGCGC
GTGCTATTATTCTTGAAGATGATGCGATTGTATCGCACGAATTCGAAGCAATTGTAAAAGACAGTTTGAAGA
AAGTTTCAAAAAAATGTTGAAATTTTATTTTATGATCATGGTAAAGCAAAAAGTTATTGCTGGAAAAAAACAC
TTGTCAAAAATTACCGTTTAGTTCACTATCGTAAACCCCTCTAAAACGTCTAAACGTGCAATCATGTGTACAA
CAGCTTATTTAATTACTTTATCTGGCGCTCAAAAACTCCTACAAATAGCCTATCCTATCCGTATGCCTGCTG
ACTACTTAACTGGTGCTTTACAATTAACTGGACTAAAGGCTTATGGTGTTGAACCACCTTGTGTATTTAAAG
GCGCAATTTCAGAAATTGATGCAATGGAGCAACGCTAACAATGAAATTAAAAAATAAATTACAAATGTTAAG
GTTGGGTCTAGGCAAATATTTCCTTGATAAAAAAAACGGATTAAACAGAATAACAAATGTTCCTAGAAGCAT
CCTCTTCCTCCGCCAAGACGGAAAAATTGGGGATTATGTGGTGAGCTCATTTGTATTCCGTGAGATAAAAAA
ATTTAATCCCCACATTAAAATTGGTGTAATTTGTACCAAACAAAATGCTTATCTTTTTAAACAAAATCCATA
TATCGATCAACTTTACTATGTAAAAAAGAAAAGTATTTTGGATTACATCAAATGTGGTCTAGCAATTCAAAA
AGAACAATATGATTTAGTGATTGATCCGACGATTATGATTCGTAATCGCGATCTTTTACTTTTACGCTTAAT
CAATGCCAAGCATTATATTGGCTACCAAAAAGCCAATTATGGTTTATTTAATATTAATCTGGAGGGACAATT
TCACTTTTCGGAACTCTATAAACTCGCCTTAGAAAAAGTGAATATTACGGTACAAGATATAAGCTATGACAT
CCCATTTGATAAGCAAAGTGCGGTCGAAATTTCTGAATTTTTGCAGAAAAACCAACTAGAAAAGTATATTGC
TATTAATTTTTATGGTGCTGCAAGAATCAAAAAAGTAAACAATGACAACATCAAAAAATATTTAGATTATCT
CACGCAAGTCCGCGGAGGAAAAAAGCTGGTGCTATTAAGCTATCCTGAAGTAACAGAGAAATTAACACAATT
GTCAGCCGATTATCCGCATATTTTTGTCCATCCAACAACCAAGATCTTTCATACCATTGAATTGATTCGCCA
CTGTGATCAATTAATCTCTACAGACACGTCTACTGTACATATTGCTTCAGGTTTTAATAAACCAATTATTGG
TATTTATAAAGAAGATCCTATTGCGTTTACACATTGGCAACCCAGAAGTCGGGCAGAAACGCACATACTTTT
CTATAAAGAAAATATTAATGAGCTCTCACCTGAACAAATTGACCCTGCATGGCTTGTCAAATAGTCTTATCT
CTTCTGACACTTGGGGCAATAGAAACTATTTCGTTGCCCTATCACTAAACTTTCTATTTTTGTGCCACATGT
TGGACAAGGCTTATCCTTATTACCATAAACCCGCAATTCTTGGACAAAATAGCCTGGACGCCCATCCGGTTG
GAGAAAATCTTTTAGCGTCGTACCACCTTGTTGGATTGCGTTAGACAGCACTTGTTTTATTTGTTCTACTAA
CTGCCCACATTGTGCCTTAGTTAAACTCCCTGCTGTTTTTTGCGGATGTAGGTTACAAAGAAATAACGTTTC
ATTCGCATAGATATTCCCAACGCCAACGACGACAGCATTATCCATTAAAAAAGTTTTAAGTGCGGTCTGTTT
TTTACGACTTTTTTGCCACAAGTAATCAGAATCAAATTCCTCAGACAGAGGCTCTGGGCCTAATTTCAGAAA
AAGAGGAAATTCGTTCAACTTCTCTGTCCATAACCACGCTCCAAAACGACGAGGATCGTTATAACGCACAAC
TTTTCCGTTATTCACTACGATATCAAGATGATCATGTTTATCAATAAGATCCCCTTTCTCCACAACTCTCAA
TGACCCTGACATCCCTAAATGTCCAATCATATAGCCTGTTTCAAGTTGGATAATTAAATACTTCGCACGGCG
ACTTAATGCGATGACTTTTTGTTGTGTAATTTGCGCTAATTCTTCGCTTACCATCCAGCGTAATTTCGGTTG
GCGAACAACAATTTTTTTCAATGATAGCCCCTTCAAGATAAGGGCTAATTCCATTTTTTGTGGTTTCAACTTC
AGGTAATTCTGGCATAGGTTATATATCCATAAATCTTATAATTGATAATATCCAAACTATTCATCAGCTATG
ATTGGCAGGCAAAAAGCCGCAATCGCGTAAATATTTTTGTCCGCAAGTCAAACAAAGCAAGGAGTCCACAAG
GCGTAATGCTTCCGCAGTAAAAGCTGCTAATGTATAGTTCGCCCTCACATTATACTCATCAGGAATATCCAA
AACACAAATATCAGAATGCTGACGCAAAGATTGATGATTACTCGCATAACCAATGAAAAGATCAGCATAATT
CTGCTCAAAAAGCCACTCTGCGGTATTTCGTCCTGTTGGAATAGTGATAGAATCCGGACCACCAACTATTGC
CATTGCTTTTTCTTTTAATTCCGAGCCATAGCCCATATGCCGTTTTTCAATATTCGAAAATAATGCCAAAGT
ATAATCTCCACAAGGATCTGCCTTAGGTGTCGATACTCCTAAGCGTAAGTGGGGCGACATCAATAATGTCAA
CCAATTCTCATCATGGTGAGTAATCACCGATTTCTTTGCAATTAAACATAAACGATTTGTAGCAAAAGGCAC
AAGTTGAATATGAGGATATCGCGCTTGTAAATGCCTAAGATGCGCATCATTGGCAGAGGCAAACAAATCCAC
TTTTTCCCCTTGCTCAATGCGTTGGCACAACAACCCCGCCGGTCCAAATTCAATTTCGACTTGTAGGTGATA
CTGTTGGATTAATGCTTGTTGCCATAACGTAAAAGGCTGGCGTAAACTCCCTGCGGCTAAAATTCTCATGCG
ATATGTTTACTGTATGGTAAAGATGGGGACTAAAACCTGCTGTTCTTCAATCATAGAATATTTAATCGGTAC
ATTATACGCTTGTTTCAAATGAGATTCCGTTAAAATTTGACTGGCTATTCCATATTTCCATTGTTGGTTAGG

CAATAGCAATAACACATTATCTGCCACACATAAACTGTGATAAGGATCATGAGTGGAAAAAATAATGGTCAT
TTTTTGTTCCGTTGCAAGAAAACGTATAAGTTGTAAGACACGCTATTGATTATAAACATCCAATGCTGCTGT
AGGTTCATCTAAAATGAGGACCTGACATTCTGTCGCAAGTGCACGAGCGATGAGCACAAGTTGGCGTTGACC
GCCCGAAAGCATATTGATATTGCGCTCAGCTAAATGCAGGATGTCTAAGCACGCCAACATCTGTAATGCGAC
TGTTTCATCCGTTTTACTTGGTAAGTTAAATGCTCCAATTTTGCTTGCTCGCCCCATTAAAACAATCTCTAA
CACGGGATAATCTGGCGACGAAAAAGACTGTGGCACAAAACCAATATGACCTTGTTGCCTAATCTGTCCAGA
CATAACAGGTAACACATGAGCAAGAGAATGCAATAATGTGGTTTTACCTTTTCCATTTGTTCCAAATACCGA
AATAACCTCTCCTTTCTTACATTGGAAAGTAAGTGGTAAATACAACGGCTTATCATAACCAAATAACAGCTT
ATCTGCATCTAAACTCAATTCATTCATAATGACTTCTTTCGATAAGTTTTTAATAGGAGCAAGGTAAAAATG
GGTGCTCCTAAAAGAGCGGTGATAATACCTACAGGAATTTCTGCAGAAGTTAACGTACGTGCAAGTGTATCA
ATAACAATCATGAAAATCCCACCAATCAAAAAGGAGGCGGCAATAGATAACGGTGATCACTTCCTACAAAA
AAACGTGTCAAATGAGGAATAACAAGCCCTATCCACCCAATACTCCCACTAACAGCGACTTGTGTTGCTACA
AGCAATGCACAAAGTAGCAAAACAAACCAACGCATTTTCTTAATGGAAACGCCTAACATTTTTGCTTGCATA
TCACCTAGCGATAACACATTAATATGCCACCGTAAACGGAATAATAAATAAGCTGCAATAAAAACGCAGGGT
AACAATATAGCTAGTTTTGCCCAACTAGTGGTGGCAAAACTTCCTAATAACCAAAATACAATGCTCGGCAGA
ACTTCTTCTGCATCCGCTAAATATTGGATTAAGCTCACTAGAGTGCTAAAGAAACCACTTAAAATGACACCC
GCTAAAACTAATACAATACGATTGCCTTTTCCGATGAACATTGTGGTTACATAGATCAAGAATAATGTCAAT
AAACCAAAAGAAAATGTGGATAGAATCAATAAATAAGATGGGAATCCTAATAAAATTGCTAAACTGCCTCCA
AAAACTGCCCCTGATGTGACACCAATAATATGAGGATCAACAAGGGGATTATGAAAAACGCCCTGTAGTGTT
GCACCACTCATCGCTCAGATCCCCCCTGAAAAAAATGCCATAATGATGCGTGGTAAGCGTACATGCCAAACA
ATATGGTATTCCATAGGTGTAAAAGACGCGTGTTGCGAAAGAAAAGGCTTAGATAAAATGGACATCACTTTT
CCGGTTGATAACGAAAAAGTGCCAATATTTAAAGTGAACAATACGATGATAAACAAGATAAAAATCAGCGAT
GTTATAAAACCTCGCTGATTTGCTAACATAGACTTCATCGTTATTACTGGTTATATGGCATACGATAGAACA
ATTTATAGTATTGGTTTACTTTTTCCTCTAAATCAACATCTGCAAACAATTCAGGGTAAAGTTGTTTTGCTA
ACCATAATTCACCAATCGCTAATGCTTCAGGCATTGGATATCCCCACGCTTTTGCATATTCCGGCATTAAAT
AGATACGTTGATTTTTCACCGCATCAATAATTTGCCAAGAGGGATCCTTTTTAATTTGCTCGATAACCTGAG
GATAACGTTCCTGTACGAAGATAACTGCAGGATTCCAATGAATCACTTGCTCAATCGAAACTTGTTTAAAAC
CTTTTATTGTTTCAGCTGCCACATTCTTCGCTCCAGCATGAAGCATCATTAACCCTGTATATTTTCCAGAAC
CATAAGTCGCTAAATCTGGATTTGCAATATAGACCCTAACACGCTGCTCATCAGGCACCTTACTTAAACGTT
GACTCACTAATTCACGCTGTTCAAAAGTGTAAGTAACTAGCTTTTGGGCTTGCGCTTGTCGATTAATTACTT
CACCAATTAAATAAATGCCTTGTTTCAAACCATTATTATAGGCAACTTCTTCATCTTCCATTTCTGGGTTGA
CTTTTCCTTCTTCACCTTTTTTATCTTCACGCAAAGAAATGGCTACAACAGGCACACCAGCCTGTTCGATTT
GCTCAATCATTTCTTTTGGTGCATAGTTTTTCCCTAATTGTTTTTTCCAACTTGATAACACTCCGACTACAC
TTTCCTTTGCATCAAGCTGGGCAAGGAGATTTAAAGTCTGATGCTGTCAGACAACAACACGATTAACTTCAT
CTGGGATAGTGACCTTTCGTCCTAATTGATCAGTAATAACACGTGCTGCAAACGCATTATTAATAGAACCTA
AGAAAAGTAATAAAGCAATACTGACTATTTTAACGTAGCGTTGAATCATAAGAGTCCCTTAATATCATTATA
TAAATAAATATATAATACTCTTATTTAGCTCATAAAGTAAACAGAAAACAAATTTGTCGTCATGAACAGAGC
GATAAAAAGGGCGTACATCACGCCCTTAATCACTTAGTTTAAAGATTATTTTCTTAATGCTTTTTTCAATTC
AGCCAATTCTTTTTGCATTGCCGATATTTCTTGTCGCAGTTGCAAAACTTCCGCAGAATTGACCGCACTTTG
TGTTGAAACCGCAGGTTTGGATTTGCTGCCGAATTTCCAAGAAACACCTGCGCCAAAGGTTTTTTCCGAACC
AGAAAAACTCCCCGCTACATTAAGCAATACGTTTTCAGCTGGCTTAAACACAGCCCCCATTGCCATCGCCTG
CGCATTTTTATAACTACCAACGCCCAAAGATAATGCAAATTTATCATCTTCGCCTAATTGTGCAGGTTTTAA
TGAAGCCAACGCCGCAGCACTTGCGCCAAGGCGGTTAATACGTAAATCTGTTCGATTTAAACGGGTATCAAC
TTGTGTAAATTGATTATTCACTTGACCTATTTTAGCATCTAAACCTTGGCCTGTTTGTAACTGCCAAGTTTT
ATCAGAACTATCTGCCACAAAAGATTGAACAGAATAAAAAGAAGTGGAAATAAGTAGACTAATTAAAGAAAG
GCGGATTAAACTATTTTGCTTGCTTAATGATTTTCATAATATTGTTCCTTTTGTCATGAATAATAATTAAGG
GTTTGAAACTTTAACAAAAAATAAAAAAGAAAAATAGGTGTTTATTTGCACATTGAAAAAGTTCATTGGTTT
TACTGATAAATAAATCTCCCCCGTCTTGCATTATCCTCCTTACAGTGTCAAACTCTCCGCACTTTTTAAAAC
TGTAAAAAATAATGACAAAAAAACGTAAAAACTTAATAAA

SEQ ID NO:76 polynucleotide sequence comprising orfs9, 10, 11, 12, 13 and non-coding flanking regions of these polynucleotide sequences.

CCGCACGCTTTCTTCTCTATAAGATCCTACAATCATAACTAATAACAATTAGCTTCCTTTAATAAAAGAAAA
AATTGAATGCCCATTAAAAATAAGCAACAATACCCAAAAAATTTCATAATATTAAGTGGGAACAAATATGGA
GCATTCTGTTCATAACAAACTGGTTTCTTTTATTTGGAGTATTGCAGACGATTGTCTGCGCGATGTGTATGT
GCGCGGTAAATATCGTGATGTGATTTTACCGATGTTTGTGCTTCGTCGTTTGGATACTTTACTTGAGCCAAG
CAAAGATGCCGTATTGGAAGAAATGCGTTTTCAAAAAGAAGAATTGGCATTCACCGAATTGGATGACCTTCC
CCTTAAAAAAATTACCGGTCATGTTTTTTATAACACCTCAAAATGGACATTAAAATCCCTCTATCAAACCGC
CAGCAATACGCCGCAGTATATGCTGGCCAATTTTGAAGAATATCTTGATGGTTTCAGCACCAACATTCATGA
AATCATCAACTGCTTCAAGCTGCGTGAACAAATCCGCCATATGTCCCATAAAAATGTTTTGCTGAGCGTGTT
GGAAAAATTTGTATCGCCCTATATCAATCTTACCCCTAAAGAACAACAAGACCCTGAGGGCAACAAATTACC

AGCGCTGACCAATCTGGGCATGGGCTATGTATTTGAAGAACTGATTCGTAAATTTAACGAAGAAAATAACGA
AGAAGCTGGCGAACACTTTACCCCACGCGAAGTGATCGAGCTGATGACGCATTTAGTCTTTGATCCGCTCAA
AGACCAAATTCCGGCCATTATTACGATTTACGACCCAGCTTGCGGCAGCGGTGGCATGCTGACCGAGTCGCA
AAACTTTATTGAGCAAAAATATCCGCTATCTGAATCACAAGGCGAGCGTTCCATCTTTTTGTTTGGTAAAGA
AACCAATGATGAAACCTATGCCATTTGTAAATCTGACATGATGATTAAAGGTGATAATCCCGAAAACATCAA
AGTCGGCTCAACCCTTGCTACAGATAGCTTCCAAGGTAATCACTTTGACTTTATGCTTTCCAACCCGCCATA
TGGCAAAAGCTGGAGCAAAGATCAAGCCTATATCAAAGACGGCAATGAGGTTATCGACAGTCGCTTTAAAGT
TACCTTACCAGATTACTGGGGCAATGTAGAAACCCTTGATGCTACCCCACGCTCCAGCGATGGACAGCTGCT
ATTCCTAATGGAAATGGTCAGCAAAATGAAATCGCCGAATGACAACAAAATCGGCAGCCGAGTGGCCTCCGT
GCATAACGGCTCAAGCCTGTTTACCGGCGATGCAGGTTCAGGAGAAAGCAACATTCGTCGCCATATTATTGA
AAAAGATTTGCTCGAAGCCATCGTACAGCTGCCTAACAACCTGTTTTATAACACAGGTATTACCACTTATAT
TTGGTTGCTGTCCAACAACAAACCTGAAGCACGCAAAGGCAAAGTTCAGCTCATTGATGCCAGCCTCTTATT
CCGCAAATTGCGTAAAAACCTTGGCGATAAAAACTGCGAATTTGTACCTGAACATATCGCCGAAATTACCCA
AAACTATCTTGATTTCACTGCCAAAGCGCGCGAAACCGACAGCCAAAATGAAGCAGTCGGCCTGGCTTCGCA
GATTTTTGACAATCAAGATTTCGGCTATTACAAAGTCACCATCGAACGCCCGGATCGCCGTTCTGCCCAATT
TACCGCCGAAAATATCTCGCCTTTACGGTTTGACAAGGCTTTGTTTGAGCCGATGCAATATCTTTATCGGCA
ATATGGCGAACAAATTTACAACGCCGGATTTTTAGCCCAAACCGAGCAAGAAATTACCGCTTGGTGCGAAGC
GCAGGGCATAGCCTTAAACAACAAAAACAAGACCAAGCTGCTGGACGTCAAAACCTGGGAAAAAGCCGCCGC
ACTTTTTCAGACGGCATCAACCTTGCTCGAACATTTCGGCGAACAACAATTTGACGATTTCAACCAATTCAA
ACAAGCCGTGGAATGCCGTCTGAAAGCCGAAAAAATCCCCCTTTCTGCCACAGAGAAAAAGGCCGTTTTCAA
TGCCGTAAGTTGGTACGACGAAAATTCAGCCAAAGTGATTGCCAAAACACTCAAGCTCAAACCAAACGAATT
GGACGCCCTTTGCCAACGCTACCAATGCCAAGCCGACGAGCTGGCAGACTTTGGCTATTACGCCACCGGCAA
AGCAGGCGAATATATCCTATATGAAACGAGCAGCGACTTGCGCGACAGCGAATCCATACCGCTCAAACAAAA
TATCCACGACTATTTCAAAGCCGAAGTGCAAGCGCACATCAGCGAAGCATGGCTGAATATGGAAAGCGTAAA
AATCGGCTATGAAATCAGCTTCAACAAATACTTCTACCGCCACAAACCATTACGCAGCCTTGCAGAAGTTGC
CCAAGATATTTTGGCGTTAGAAAAACAGGCTGACGGCTTGATTAGTGAAATTCTAGAGGCTTAATAAAAAAC
AAACTATTAAGCAAGTTTTAATAGGTCTTAAGTAAGGAAATTCAAAATATATAACACATTGAAAAATAATGA
ATTTTACCTTTTAAGCAAGATTTGGCATGAAATAAGCAAGGAATAATAATGACAGAACCGCTTTCTAAAATT
AACGGCATTATCACAAAAAATTATTTAGAGATGCAGCCGGAAAACCAATATTTTGAGCGCAAAGGACTAGGA
GAAAAAGACATCAAGCCAACTAAAATAGCTGAAGAATTAGTTGGAATGCTCAATGCTGATGGCGGAGTTTTG
GCTTTTGGTGTGGCAGATAATGGCGAAATCCAAGACTTGAATAGCCTTGGCGATAAATTAGATGATTATCGG
AAATTGGTTTTCGATTTTATTGCACCGCCTTGTCGGATTGGACTGGAAGAAATTCTGGTTGATGGAAAATTA
GTTTTCTTATTCCACGTAGAGCAAGATTTAGAGCGTATTTATTGTCGCAAAGACAATGAAAATGTGTTCTTA
CGTGTAGCAGATAGTAATCGAGGCCCTCTCACCAGAGAACAAATCAAAAATCTTGAATATGATAAAAAATATC
CGTCTATTTGAAGATGAAATAGTTCCTGATTTTAATGAAGAAGATTTAGATCAAGAATTATTAGAGCTATAT
AAAAAGAAAGTTAATTTTACCTCCGATAATATCTTAGATTTATTATACAAGCGAAATTTATTAACCAAAAAG
GAAGGTTGTTATCAGTTTAAAAAATCAGCCATTTTACTCTTTTCTACCATGCCGGAACGTTACATTCCTTCA
GCATCAGTCCGCTATGTTCGTTATGAAGGTACAGTAGCGAAAGTCGGTACTGAGCATAATGTGATAAAAGAC
CAACGTTTTGAAAATAATATTCCAAAGCTAATTGAGGAGCTGACCTATTTTTTAAGAGCCTCTTTAAGGGAT
TATTACTTTCTTGATGTCAATCAGGGAAAATTTATCAAAGTACCGGAATATCCTGAAGAAGCTTGGTTAGAA
GGTGTTGTAAATGCGCTTTGTCATCGTTCTTACAATGTTCAAGGTAATGTTATTTATATTAAACATTTCGAC
GATCGTCTTGAAATTAGTAATAGTGGCCCTCTCCCTGCTCAAGTCACCATTGAAAATATTAAAACGGAACGA
TTCGCTCGGAATCCACGTATAGCACGAGTTTTAGAGGATCTTGGGTATGTCCGTCAGCTTAATGAAGGCGTT
TCCCGTATTTATGAGTCAATGGAAAAATCATTATTGGCAAAGCCTGAATATAGAGAACAAAACAACAATGTT
TATCTAACATTGCGCAACCGTGTTACCGCACATGAAAAAACGGTATCTACAGCCACTATGCTGCAGATTGAA
AAAGAATGGACAAACTACAACGACACCCAAAAAGCCATTTTGCTTTATCTATTTACAAATGGTACGGCGATA
TTGTCAGAATTAGTTGACTATACAAAAATCAATCAGAATTCGATCCGAGCGTATTTAAATGCCTTTATTCAG
CAAGGTATTATTGAAAGACAAAGTGTAAAACAGCGTGACCCCAATGCCAAATATGCTTTTAGAAAAGATTAA
GCAAGGTTTATCGCTTGCTAAGCAAGGAAATTGACAATGCTTAACTTGCTGAAAAATAATGATTTTTATCTT
TTAAGCAAGATTTGGCATGAAATAAGCAAGTTTTTTATAGTTAAACGGACAACAAATTGCATCAATAAGAG
CGGTCATATTTTAAGGATTTTTTGCAAATGAGACGATACGAGCGTTACAAAGATTCAGGTGTGGATTGGCTA
GGGGAGGTACCGAGCCATTGGGAGTTAAAACGCTTGAAACAATTATTTGTTGAAAAAAAACATAAGCAAAGC
CTGTCTCTTAATTGTGGAGCCATTAGTTTTGGTAAAGTTATTGAAAAATCGGATGATAAAGTAACAGAGGCA
ACAAAACGTTCATATCAAGAGGTGTTAAAAGGCGAGTTTTTAATAAATCCTTTAAACTTAAATTATGACCTA
ATTAGTTTGAGAATTGCTTTATCAGAAATAGACGTTGTTGTAAGTGCCGGTTACATTGTTTTAAAAGAAAAA
CAAATAATTAATAAAAAATACTTTTCGTATTTATTACATAGATACGATGTTGCATATATGAAATTATTAGGT
TCAGGTGTAAGACAAACGATTAACTATGGGCATATTTCAGACAGTATTTTGGTTATTCCACCTCTCTCCGAA
CAACAAAAAATCGCGCAATTCCTAGACGATAAAACCGCTAAAATCGATCAGGCGGTGGATTTGGCGGAAAAG
CAGATTGCCCTGTTGAAAGAGCACAAGCAGATCCTGATTCAAAATGCCGTAACCCGAGGCTTAAACCCTGAT
GTGCCGTTAAAAGATTCCGGCGTGGAATGGATAGGGCAAGTGCCGGAGCATTGGGATGTGCAACGTTCAAAA
TTCATTTTCAAGAAAATAGAAAGAAAAGTGAATGAGGAAGACCAAATTGTTACTTGTTTTAGGGATGGGCAA
GTAACTCTGAGAGCTAATCGAAGAACTGAAGGATTTACAAATGCGCTAAAAGAACACGGCTACCAAGGAATT

AGAAAAGGTGATTTAGTTATTCACGCTATGGATGCTTTTGCAGGGGCAATTGGTATTTCTGATTCAGATGGT
AAAGCAACACCAGTTTATTCCGTTTGTTTGCCTCATGATAAACAAAAAATCGATGTCTATTTTTACGCTTAT
TACTTAAGAAATCTTGCATTATCAGGATTTATTAGCTCCTTAGCTAAAGGAATTAGAGAGCGTTCAACAGAT
TTTCGCTATTCTGATTTTGCAGAATTATTACTACCTATTCCTCCATATTTAGAACAGCAAAAAATTGCCGAC
TACCTAGATAAACAAACCTCTAAAATTGATCGAGCAATCGCATTAAAAACAGCCCATATTGAAAAGCTGAAA
GAATATAAAAGCGTGTTGATTAACGATGTGGTGACCGGCAAGGTGCGGGTATAGGTGTGAAAAGTGCGGTCA
AAAAATCCGATGGATTTTGAATATCGGCGCGACAACTTGGGCGTAATGAATAAATTTAAAAAATTCACAAAA
GGGTGAAAAATGGTTTCAGGAACTAAGGAAAAAGATTTAGAAATTGCCATCGAAAAAGCCTTAACTGGCACT
TGGCGTGAAAACATGGAAAATAAGCTGGGCGAGCCGAAGGCTGAATACCTGCCGCGCCATCATGGTTTTAAA
CTGGCATTTTCACAGGATTTTGATGCGCAGTTTGCCATCGACACACGTCTGTTTTGGCAATTCCTGCAAACC
AGCCAAGAGGCAGAACTTGCCCGTTTTCAACAACTCAACCCAAACGACTGGCAGCGTAAAATTTTGGAGCGA
TTAGACCGCCAAATAAAGAAAAACGGCGTGTTGCACCTGCTGAAAAAAGGCTTGGATATTGATAGCGCCCAT
TTTGATTTGCTCTACCCCGTTCCGCTTGCCAGCAGCGGCGAAAAGGTCAAGCAGCGTTTTGAACAGAATTTG
TTTAGCTGTATGCGTCAAGTGCCTTATTCTGCCTCAAGCAATGAAACGGTGGATATGGTGCTGTTTGCCAAT
GGCTTGCCGATTATTGCCCTTGAGCTGAAAAACCATTGGACAGGTCAGACAGCCATTGATGCGCAAAAACAA
TACCTCAACCGTGATTTAAGCCAAACGTTGTTCCATTTCGGGCGTTGTTTGGCGCATTTTGCCTTAGATACG
GAAGAAGCTTATATGACCACCAAATTGGCGGGGCCTGCTACGTTTTTCTTGCCGTTTAACTTGGGCAACAAC
TGCGGTAAGGGTAATCCGCCCAATCCCAATGGACACCGCACGGCGTATTTATGGCAAGAGGTGTTCGGCAAA
GCAAGCCTTGCCAACATTATTCAGCATTTTATGCGCTTAGACGGTTCAACCAAAGATCCGTTGGATAAACGT
ACCCTCTTTTTCCCTCGCTATCACCAATTAGATGTGGTCCGCCGTTTGATTGCTGATGTCAGTGAACATGGC
GTGGGTAAACGTTATTTGATTCAACATTCTGCCGGTTCGGGCAAGTCTAATTCCATTACTTGGCTGGCGTAT
CAGTTGATTGAGGCATATCCGCGCAATGAAAAGGCGGCAAACGGTAGAGAGGCAGACCGCCCGATTTTTGAT
TCGGTGATTGTCGTAACCGACCGTCGTTTGTTGGATAAGCAACTGCGCGACAATATCAAAGATTTTTCAGAA
GTTAAAAACATTGTTGCGCCGGCGTTGAGTTCGGCAGAGTTGCGCCAATCGCTTGAGCAGGGCAAAAAAATC
ATTATTACCACGATTCAAAAATTCCCGTTTATTGTCGATGGCATTGCTGATTTAGGCGACAAACAATTTGCG
GTGATTATTGATGAGGCACACAGCTCACAATCAGGTTCGGCACACGACAATATGAACCGGGCCATCGGCAAA
ACGGAAGACCTTGATGCTGAAGATGTGCAAGATTTGATTTTACAAACCATGCAATCCCGCAAAATGCACGGC
AATGCGTCGTATTTTGCTTTCACCGCCACACCGAAAAACAGCACTTTGGAAAAATTCGGCGAAAAACAGGCG
GATGGCAAGTTTAAGCCGTTCCACCTTTATTCTATGAAGCAGGCGATTGAAGAAGGCTTTATTTTGGATGTA
ATCGCCAATTACACCACCTATAAAAGTTTTATGAGATCACTAAGTCGATTGAAGATAATCCGGAGTTTGAT
AGTAAAAAGGCTCAAAGCCGTCTGAAAGCCTATGTGGAGCGTTCGCAACAAACGATTGATACTAAAGCGGAG
ATAATGCTGGATCATTTTATTTACCAAGTTTTCAACCGTAAAAAACTCAAAGGCAAAGCCAAGGGAATGGTG
GTAACGCAAAATATTGAAACCGCCATCCGCTATTTTCAGGCGTTAAAACATTTGCTGGCCGGGCGGGGTAAT
CCGTTTAAAATTGCGATTGCGTTTTCAGGCAGTAAAGTGGTTGACGGTGTCGAATACACCGAAGCGGAAATG
AACGGCTTTGCAGAAAGCGAAACCAAAGAGTATTTCGATCAAGATGAATATCGTTTGCTGGTGGTCGCCAAT
AAATATCTGACCGGTTTCGATCAGCCGAAATTGTGTGCCATGTATGTGGATAAGAAACTCTCCGGCGTGCTT
TGCGTGCAGGCTTTATCTCGTTTGAATCGCAGTGCGAATAAGTTGAGTAAACGCACGGAAGATTTGTTTGTA
TTGGACTTTTTTAACAGCGTTGAAGATATTCAGCAGGCATTTGAGCCGTTTTATACTTCTACTTCGTTGTCG
CAGGCAACCGATGTCAATGTCTTGCATGATTTGAAAGACCGGTTGGATGAAACCGGCGTGTACGAACAAGCG
GAGGTCAACGATTTTACTGAAGGCTATTTTGCCAATAAAGACGCACAGCAATTAAGCAGTATGATTGATGTG
GCTGTCCAACGTTTTGATGATGAATTGGAATTGGATTTGGATCGAAATGAAAAAGTTGATTTTAAAATCAAG
GCAAAACAGTTTTTAAAAATTTACGGGCAAATGGCCTCCATCATCAATTTTGAAAATATCGCTTGGGAAAAG
CTCTATTGGTTCCTCAAATTCTTAGTACCCAAATTAAAAGTACAAGACCCGATGGATGAATTTGATGAAATT
TTAGATGCAGTGGATTTAAGCTCTTACGGCTTGGCGCACACCAAGCTGAATTACAGCATTAAATTAGATGAT
GAAGAAACAGAGCTTGACCCGCAAAACCCCAATCCGCGCGGTACGCATGGTGAAGATAAAGAAAAAGATCCG
ATTGATGAAATTATTCGTGTATTTAACGAAAGATGGTTTCAAGATTGGAGCGCAACGCCGGATGAGCAACGG
GTAAAATTTATCAATATTACCGAGCGCATCCGCAGCCATAAAGACTTTGAGCAGAAATATCAAAATAACCCG
GATATTCATACCCGTGAATTGGCTTTCCAAGCCATTTTGCGCGATGTGATGAGCGAACGCCATAGGGATGAA
TTAGAGCTATACAAACTTTTTGCCAAAGATGCCGCATTTAGAACCGCTTGGACGCAAAGTTTGCAACGGGCT
TTGGCTGGATAGAAAAGATTGCCTGAAAAATTAACGTTCGGCTCTCCTTTTCTATCTAAATTAATATCATCG
TAAACATTAATTAATTTTTTCACATACTTAAAAGAGAAAATTAAATATAGTTTCCATAACAGCAACGTCGTT
AATTAGAATAATTTATAAATTAGCTATAATT

SEQ ID NO:77 polynucleotide sequence comprising orfs14, 15, 16, 17, 18, 19, 20, 21, 22 and non-coding flanking regions of these polynucleotide sequences.

TTGATTTACACGATCAGAGTTTGGATCTTTGATAATCATCGGAATGTTGTATGGCTGTTTAGAACCCTATCC
GCCTTGTCGTTGCAGAAAACGCTGGTTTCACTTCACATTCCCCTTGTAGTGCCATCAGCCAATCTTGCACCT
TGTGAAAAGGGGAAAATTGGTGACGACGAGCCACTGCAGCAAAATCCCCTGGTGTCAGCAGATTAAGCGATT
CAATCTGACTTAAATCCTCTTCCGATAACAACGGCAATCCTAAAATTTCTGCTTGTTGTTTAGCAAAATCTA
AGCGTTGTTTGAGCGTTAAATAATCAAACTTCAATTTTAAATCAAAACGGCGTAAAGCTGCGTGATCAAGAA
CCTCAATTAAATTTGTTGATACCACCATCAGGCCCTCAAAGCGTTCAATTTGTGTTAGCATTTCATTCACTT

GCGAACGCTCCCAGCTTCGATTTGCGCCTTCTCTAGAAAATAAGAACGTATCTACTTCATCTAGCACCAATA
TTGCATTATCGGCTTTCGCTTGTTCAAAGGCTTGAGCAATATTTGTTCTGTCCCGCCCACATAAGGATTAA
GTAAATCTGAGCCTTGTCTTAGCAATAGCGGCATGTCCAACTGTTCCGCAAGCCACGCTGCCCAAGCAGTTT
TTCCTGTTCCCGGCGGGCCATAGCAACAAATTCGCCCTTTTTTCGACCGTTTTAACCCTTCACTAATACGAT
GAATATTGTCGTTACAAGCCACATAATCCAAGTTGTAGTCGGCTTTGCCTAAAACAAGCGGTTCAATTTTCG
GTTTATTTTGCGATTTTAACGTTTGATTAAACATCATGAGCAAAGTCTCAGCAAAATTTGATGTATTGAGTT
CCTTTGCCACCCGAATTGTGCGGCTTAAAATCGCCGGCGTTAATGACCGCACTTTAGCAAAATGCTGCACAT
AGGCCGGACTTAATTTTCCCTCAGTCAGTTGCGTAATCAGTGCTGACTTATTTTTCAACGGCAAATCTGGCA
TTTCTAAAATAAAATCAAAGCGGCGTAAAAAAGCAGGATCTATGCCCGAAACAGAGTTAGATAACCAAATCA
TCGGCACGTTATTGTTTTCCAATAACTGATTTGTCCACGCTTTATTTTTTTGTGCAACAGAACGCTCCATAA
ACGAGCCGTTAAACACATCTTCAATTTCATCAAAAATTAAAAGCGCCTGCTTGCCGTTCAATAGCGTTTGAG
CAAGACGACTGTAGTTCAGGCGTTGCTCTGCCTCCACAACATCTCCGTCAGAATCCATGTAAGTAATGTTAT
ACGCCGAAATCCCCAACGCCTGTGCAAGCAACCCGGCGAATTCTGTTTTACCAGTGCCAGGCACGCCATAAA
TTAAAAGATTCACGCCTTTTCGATGATGTTTTAGTGCTTGTTGCAAATAAGTCAACATCATCTCTTTCATGC
CGGCAATATGGTCAAAATCATCCAGTTGCAGACTTGGCACTTGAGCGACTTCCGTACAAGATTTTAATAGGA
CGTTTTCGTTTAATGGTTGTGTCACAAATTCATCAAAATCTAAGGTTTCGCCCCAATCTAAATAATCATGCA
CACTATCGGGGCGATAATCGCGATCAATCAGGCCATAAGCATCGAGTTTACTGCCTTTCTTTAAGGCAGATA
GAATCTGATTTTTCGGCTGTTTAAGTAAATCCGCCATGATCGCAGCCGTTCTTTGTAAATCCGATTTCGGCA
AGTAGCCAAACAAATCTCGCATAGCTCCTTCACTACGTAAATGCATGGCAAAGCGGAGAAGTTCCTGTTCAA
CGGGATTCAGTTGCAAAAATTCTGCCAACGTTGCCAAATTTTCATACGCCTGTTTCCATAACTCAGGTAAAA
GTGCGGTGGATTTTTGGAGTTTTTTATACCGCTCTTTTAAAAGCCGACGAGCAACCGTGCGTAAATTTTTAT
CATTCTCTAATTCTTCAGGCAGCCCAAATGCACTGGCAATTTCATCACTTCGCCAGCTAGTCTCCCGAAACA
CTTCGGAAAAACCTTTATGCTCAAATAAAACTTTAAGCATCATATTTTCAGTATAAGAAGACACTGTCGGTG
GGTTTAATTTATATTCAGACATAAAAAAATACTCCTTACTGGGTTGGTAAGGAGTATTTTAGTGAGTAGTGC
GACAAAAGGTGTCGTTAAGGATAGTTTTAAGAACGTTTGTTAATCAACCATTCAACTAAACCAGCACTAATT
ACAAGCTCTGCCATTTTTCGGCCATTTACAAGCTTAATTCCTTTAGCTTGATATTGTTTTTCATCTATGTTT
TCTAAACCAGAATGATATACGTAATACATTTCTGAATAACCATAGTTTTTATATTCACTTTCAAAGTTCGAA
ACATATTCGTCTAATTGTTTAATATCCGTATCTGACTTAATTTGCACAAATACTCTCTTCTGCGTTGAAGAC
GAATACAAATCAAGATCTATTCCTTTCTCCGTTTTACCTAAAACAGAGTATCGTTGCCATCCTAATTTAGAA
AAAACAAGATCCGTTAAAAGTTCAAAGTCACTCCACCATAAACCTTTAATTAATTTTTCAACTGATTTAATT
AATGTTTCATACGCCTCTTTCGCTTCTGTAATTTCCTCAATAACTTCACCATTTATACGACGTATTAAATAG
TCCTCCATCTCAACACCACAAATCGTCCCTCTATAGGCTTGGACCTTTGTTACTCTACCATCAAGATTATCG
ACTAAAAGCTCTTTACCGTTAGCATCAACGCAAGACCAATTCCCATTGTTACTAATAACTTTTCTTGTTCTA
GAACCATCGCTTTCCTCAACAACCTCTTTACTGCAAAAAGCCCAATATAATTTACGTCCAAAGAAGGTGATC
CAAAGTGTATCTTCCCCAAGTTGATAAAAATCTTGAATTTGTCTCAAGTGATTTGAAACAGTTCCTGTATGG
TCACTCCAATAAGTTTTACAATATTCAATACAACTATCCCATTGATTATTCAAACATTCTTTGTGAATCTCT
GATGTAGATTCATAGCCAAGACGAATCGTATTTTTTGTACTTGCTGTACTATTTTTATCAATACAATCTTTT
TCCCAACATCCTTTTATGCCTAATTTAATAAAACGAATATTAGTAGGTTCAATTTTTTCAAACATAGTTTTT
CCTTATTTCTAGTTAAAATTCACCGAATTATAGATAATTGAGCAAAAAAAAAACAATTTAAACATATTTTTT
ACTCAATAATAGAATGACAACAAACTACCGACAAATCATCCGAAAACGATTGCTTCTCAATCATCTTGCGGC
AAACCGTAAGGCGATATTTATCATCGGGATATTTCTGCCAAATTTTTTCGCGCATTTCATCTGAAAGCCCGT
CGGTCAAGCCGTCAGAACAAAGTAATAAACTTTCCCCTTGCTGAATTTCAATTTCTTGATAAAAAATTTTAT
CTTGAAATTCGGAATAATCGGCGACTAAACAAGAAGAAACGCCGCCATAAATCGTGGCAAAATCTTCTTCTT
TTTTATCGGGGAAATCAGTCAATAATTCAGAAAGAATAGAATGATCTTGGGTGATTTGTTGCCATTTTCCTT
GGGCATCAATTAAATAAGCACGACTATCGCCTACGCTGAGAATTTTCGCTTTACGGGTTATTTGATCAATTT
CGGCAGCCACAAATGTGGTCGCCGAACCAAAATAATCCTCAGCTAATTCTGCTGATAAACTGGATTGTAAAT
CGTAGATCGTTTGACGGTTTATACTTTCCATTTGGCTTAATAATTGCATAGCCAATTTGCTCGCTTTTTCAG
GTCGGTTGCTATTAGAAATACCATCTGCCACGCCCACAATAAAGTGCGGTCGGTTTTCAAGGCGTTTTTCAG
CCGTTTTGAGTTTATATTGAAACACCGCCTCGCCATTAAAAAGGGCATCTTGGTTGCGTCGCTTGTTGCTGC
CAATTTTGTTGGCAAAGGGTAATTTCGCAAAAATTTTTCATTTATTCAACCGCTTGTTGAGAAGGATTTAAA
AGGCGATCAATCGCTTTTAGTGCATCTAACGCTTTCATTTCTTAGACTTAAAAAAGTGCATTTTCGGGCACG
CCCTGCATCTTGTGGGGTAATACGGGATAACCCCCCCCTTTTTTTTGCTTTTCGCCGTACGTTCAGAAAATC
GACGCACAGTGGAATGGCTTTTCCTGTTCCCAGTTCGATAACGACGAGATTTTGCACTTCTTTTAACCACGA
TTCTAACCGCACTTTTTTAAAATCCTGATATTGACTTGCATAACTCCAATCATTAAACATTAGTACATTTTG
ACGAGCAAAGCCCCCACAATAAGGCAAATGTGGTTTTTCACTGGTTAAACATAAGTTTTCATTATCCACGAC
AGGTTGAAAACTTGATGCAGACCAACTTAATCCTCGACAATTATTGACACATTGAAGACGCTCCAAAGTACC
ATGTACTTCATAAACATGGCTATCATTAAAACCAGCCTTTTGAAAATGCCCATCAACATTACTGGTAAAAAC
AAAATATCCATGAGGTTTATCTCCCGCCCAGCATTTTAAAATCTGATACCCTTCGTGAGGAAGAGTATTTCG
GTATTGAACTAATCGATGCCCATAAAACCAATAGGCTAGTTCCTGATTATGCTTATAAGCTAGTGGCGTTGC
GATCTCTTCAAAAGATATATTATGTTCTTTAAACATAGGATAAGCATTCCAAAATCCGCCAACGCTGCGGAA
ATCGGGAAGCCCAGAATCCACGCTCATACCCGCACCAGCTGTAATTAAAATGCCATCCGCTTTGCGGATAAG
TTCCACTGCATAATTCAAATCATTTTTCATAATACTTTTCTCTGCCCATTTTTCATTGATGAAATAATACCC

GCTTGTTCCAACTGTTCTAAAATTTGCCCAGCTCGATTAAAACCCAACATAAATCTGCGTTGAATCATTGAG
CAAGATGCAAATTTTTGTTGTTGCACATATTTTTTTACATCCTCAAAAAGTGGATCTCTTGCCATAATTACT
TTCATTGCCATTATTTGCTCCTTTTTCTTAATTAAAGGCTTTATAAATATGTAAGAAGTAAAGAATTTCTCT
TTATGGAGAAATTATATGAAAGGAAGCGACAACTTGTGTCGTTTGTGAATATTGAAAGCGGTTATTTTTAGA
AGATTTTTTGCAAATAAGATGCTCTGTATTGCAATATGCATATTTATCTGGTTATATATACATGTTAGTTAT
TAAGGAAAATAATATGAATAACCAAAACCCGATTGAAATTTACCAAACTCAAGATGGCACAACGCAAGTGGA
AGTGAGATTTGAAAATGACACCGTTTGGCTTTCCCAAGCGCAGATGGCTATGTTATTTGGTAAAGATATTCG
CACCATCAATGAGCACATTACCAATATATTTGATGACGAAGAACTTGAGAAAGAATCAACTATCCGGAAATT
CCGGATAGTTCGCCAAGAAGGTAAACGCCAAGTCAATCGTGAAATTGAGCATTATGATTTAGATATGATTAT
CTCTGTTGGCTATAGAGTAAAATCTAAACAAGGCATTAGTTTCCGCCGTTGGGCAACTGCACGTTTAAAAGA
ATATCTGACTCAAGGCTATACCATTAACCAAAAACGTTTACAGCAAAATGCTCACGAATTAGAACAAGCACT
TGCGCTTATTCAAAAAACGGCAAATTCATCGGAATTAACGCTAGAAAGCGGTCGCGGATTAGTGGATATTGT
CAGCCGTTATACGCATACGTTTTTATGGCTACAACAATATGATGAAGGTTTACTTGCCGAACCACAAACACA
GCAAGGCGGTACATTACCGACTTATGCTGAGGCTTTTTCTGCACTAGCAGAGTTAAAATCACAGCTGATGAC
AAAAGGTGAAGCAAGTGATCTCTTTGGACGTGAACGAGATAACGGCTTATCTGCGATTCTAGGTAATTTAGA
TCAAAGTGTATTTGGTGAACCTGCTTATCCAAGCATTGAAGCAAAAGCGGCGCATTTACTTTATTTTGTCGT
CAAGAATCATCCTTTTTCAGATGGTAATAAACGTAGCGGCGCATTTTTATTTGTAGATTTCTTACATAGAAA
TGGGCGTTTGTTTGATCATAATGGATACCCAGTTATCAATGATACTGGGCTTGCCGCGCTCACTTTATTAGT
TGCTGAATCTGATCCGAAACAAAAAGAAACGCTTATTAGGCTTATTATGCATATGCTTAAGCAAGAGAAAAA
ATGATAAATAGCGACCGAAGTCGCTATTTGTTTAAAAAGTGCGGTCATTTTTCTATGAGTTTTTGGTGTTCT
CTAATAACTCTGCCACCACTTTTGGCACACCCTCGCCTGCTTTTTCTTTGATTGCAATAACTTGCTTACGAA
CAAATCCTGTATTTGGGTTAGGATCAATCAGATAAATTGGCGCTTTTCTTGGGGCTTCATTGACTAAGCCAT
TGGCTGGATACACTTGTAAAGAAGTGCCAATCACTAACACAACATCTGCTTGTTCCACAATATCAACCGCTC
GTTCTAGCATCGGCACCATTTCACCAAAAAAGACGATGTAAGGGCGCATTGGGTGTCCATTTGGATCTTTAT
CTTCTAATTTCTGATCACCAAAACAATCCACAATATAACTTTCATCAAAGCTACTGCGAGCTTTATTTAATT
CACCGTGTAAATGCAACACCTTCGAGCTGCCGGCACGTTCATGTAAATCATCCACATTTTGCGTGATGATTC
TCACATCATAGGCTTTTTCTAGTTCAACTAAGGCGAGATGCGCAGCGTTTGGCTTAGCTGCTGCCGCATTTT
TACGGCGTTGGTTATAGAAATCAAGCACTTTCGCACGGTTCTTTTGCAAGGCTTCGGGCGTACAAACTTCTT
CTACTTTATGCCCTGCCCACAAACCATCTTCCGATCTAAAAGTTGGAATTCCACTTTCGGCACTAATGCCAG
CTCCCGTTAATACCACGCAAATTGGTTTATTTTTCTCTGTCATTTTTCAGGCTCCTTTTATTAGCAAACTGT
TCTGTACCAAAATGAACATGCTCGCCTTGAAATTTGCCAGCACCATTTTTAGTGCGATCATCAATTAAATAA
TCACCTTGGTTGAGATTTTTATGATGGGATAAAATCAATCGTTTATATAAGGCTGAACCTTTTTCTTCACCG
AAATAATGGTGAATCCATTTTACTTTTATACTCCCAAGCAAAAGGATTATGCCAAGGCGCAGTAGAAAGCAC
ATAAATATGATATTTTTTCATCAATTTATGCACCGCAGAAATCGCATTCGGCATAGGTTCCATTAAGCTAAA
AATGCCCTCGACTTCATCATATCGACCTTCATATTCTCGCTTGGTTTTATCATCTAGTTTTGCAATACCTGA
TGGAAAATCTACCATCACATTATCCATATCAATATAAACAATTTTCTTCATTTTAATGCCCTCTCTGTTGAT
GGCTTAATGATAAAAGATGAAGCGACAATTTATGTCGTTAGGCATTTTCGTCTAAATAAGTGCGGTCAATTT
CTTGGTAATCTTCACCAAAATGGGCTATCCACCATTCCAGCATAGCGCTCTTAATCACGGTAGCGGAAATTT
CATATTCAGTGCCACAATCTTTTACTGTTTGATCCATTGATAATGGTGTTTCTGTTAAAAATCCACCAATAT
CTTTATTAATGCGGAAAGTTAATCGAATTTTTCGACCATAGGTAAAACCAAACTTTTGGCTTTCTACATAAG
ATTTCAAATTAAAATCAGGGCGTTCAAATATCATTGTACTCACTGTTACCTTAAGCAAGCGATGCAAAGCAA
GGTGTAAAATATCGCCATTCTCATATTGTGCGACTAAATAGCTACTTGGTCCTTGTTGAACCAAAGCCAAGG
GCTTGACCTGTGCCTTATGTTCTTTACCGTGAATACTCCGATAATGCA

SEQ ID NO:78 polynucleotide sequence comprising orfs 23, 24 and non-coding flanking regions of these polynucleotide sequences.

CAGCTTAAGGGAGAACTGGCAAAGGTGAAATTAATTTCGTAATAAATCAGAGCGTATCCATCAGACTCTCAT
GTTCTGTTTGTTTAAATGTAAGTACTAACTCTTTATAAGCTTCTAGATCTTGATCAAATAATGCCCGTGAAT
ATGAAATTTTATATATTACTTCCCTTTCATATTCTTCATCAATTAATTCATTGATTCTTTCATAATCTTCAT
ATCGTATTCGTTCACTTTTTCGATAATCGTTTGTAGCAATGTAAAAGTGTAGAATAAATCCTAAAATTGCAT
TGGTTGAATGAAGTACAAATAAAGCAAGATCGCTACTTACTTGCTTATGTTCAATATCTTGACCGTGAGAAG
CATAACTACCATATTCATTTCTAATTTTTGCAACATAATGAAGAATTGAACCCAGACTTTTAGCTAATTCAA
GCAAATATTGGTAATCTTGATGATAATTCAGATCTAATTTTTTAATTGTTGTAGATACAAGATTCGGATATT
TTTCAGGAATACTTTCTCCTTTATCATTGAGAATTGTTTTGCAAATACCTTCTGTTACAGATTTGCACAATT
CAATACTTAAGATTGGATTCGTATAAACACTTCTGATGATATTATCAATATGTCCATGATAATGCTGAAAGC
TAGGTGCTTTCTCCATTGACCCAAGCACCCAGTTCATCATAATTGAATTTCTCCACTCAATAATCTAGGTAA
CAATAAATCCCTTATTTCTTTCAGTGCATTATTTTCTATCTCGTTATTCATAATTTTTGAATCACAAGATGA
TAAATATTTTTCAAAAAGCTGAATAAATTTTTCATCAGGGTTAATAATTTGGATATTTTTTAAGTTATCCTG
ATTGATAGAACCAAAAACAGTTCCTTCACCATTAAATAAATCTAATTCTGGTTTTATAGATTGTATTTGATA
TAAACCGAACGACAAACTTTTACTCTTATGTTGTAATGCAGCTAATCCGCGACCAATACAGCATTTTTCAAG
TGCTATATTAATGTCCCCAACAGGAGCTCGAACGCTCATTAAAATAGAATTTTGTTCTGCAATACGTTTAGG

ATCTGTTGTAAATAATCTTGGGGTAGGAAAGCGCCAACCAAATTCTGCACGACCTTGATAGAAAAGCATCCC
TTGTTTGTTTTCATTATAAGTTTCTCCTTTTGGAGATTGCCCCATAACGACATCATAACAATCGCCAATCGT
AGATAATTCCCACCCCTTCGGCACTTCAACCCCATCAACCTCCACCATCTCACACGGAAACGCTTTGGCGGT
TTCGGCTAGTTCGGCGTAGCGGTCAGGCTGTGTTTGTGAAAGTGCGGTCAGTTCTTCGGGTGTTTTTCCGCT
GATTGCCTGCATGGCGGCAAGTTCTGCTTGTTCAAGGCTAAGACCGTCTGAAAGGGCTTGGATTTTGGCACG
CACGGGATCGAAATCGACAAACCAGCTTTTAAACAGGGCTTGGGCGATTTGTTCTAAGGTTTGGTTGATTTG
AGTGTTGAGTTCTATTTTTTGATCTAAAGTATTTAGAATATATCCAATTTCCTTTTGCTTATTTATATCTAG
AAGTAATAATTTAACTTTACTTAAAGCTGATACATATAGATTTTTTTGGACACTACCTTCAGCTAAAGTTTC
AATTTGTTCTTTGCTATTTAATAAGGTATAAAAAATAAATAATGGGTTACAAATATTTTCATTAACTTTGAT
ATTAATTACAGCTCTATTTCCACACATGTAATCTTTTAAGATTCCAATTCGTCCAATAGTTCCTGATTTGCT
AATTGCTAAACTATCTGGTTCAAATAATACAGCACTCTTCTTTGCACTTAAAAATCCTTTTTCTGTTAAAGT
TTGAGAGGTTTTATATACAAAACCATTGTTTAAATCTGTTGCTCTCAACCATTTAATTGTTCCTCCAAAGTA
GCTTGGTTCATTTTTTGATGGATTAACATAACCTTCTTGAAATGAAGCGCAATCAGCTAAAGTTATAACCTT
CCAATCATTCAT

SEQ ID NO:79 polynucleotide sequence comprising orf25 and non-coding flanking regions of these polynucleotide sequences.

CACGCTAGTGCCGCCTCAATCCGACGCGACTGCGTCGCAATCGGTTAATCATAAGTGAGTGGCGTTGCCACT
CGTGTTGGAGAACACAGCCCCCAGCGGGGCTGAATTATGCGTAACCATGTACGGCTTTGCCGTGCATGGGAA
AAAATAAGCGGTGAAATCTTGCAAATTTTTTGCAAAATCTTACCGCTTGTTCTTTTGAAAAAAGCATTAAAA
CTCATCTAAATCATCTTCATGATTCATTGATTTTTTATGTCGGTATCCATTCTTATATTTAATTGCAAGTTC
CATATAATCTTTATTTCTAAGTTCTTCATCTTCAGCTATTTTTTCAATTAAACTATTTACTTTATCCTCATC
TCCACAAATTTTAATTAAGGCATCCCAAAGTAGAATTTTCTCTCTATGTATTGTAGGATCATCCCCTCTTTG
AGATTTACGTTCTGATATTGAAGATTTAAGTAATGATAAAAATACTTCAGGGGAACTTAATATATCATCAGA
AAAAGGGTTTCCAATTGAAACAAAGAAATAGATTATATGTGACAAATTATAGGTTCCTGCAAGTTCTTTAAT
TGTTGCAGAGCGAATATTATTTAAAAATATTTCCTCCAAGTCTTTTGCATCCGATTCAGATACTAATTGATG
ACCTCGGCCCTCTCGATATCCAATAATTCCTACAATTTGATATTGCCCATATAGATCGCTAGAATTTAATAG
TTGAGTAATAACTTCTTTTTTATCCTTCTCAGGAAGTCTTCTAAGTAATCTATAAACTAAGCGACTCCAAAC
CATATCCGCCCCAAAGTCAAAGAATCCTAATTCTTTTTCAGGCACTCTTGGTAAATTTCTATATAATGTTGG
TATAGTTGCTAGAGCTATTTCTTTAGTAAAGTCTTTTTCATAGTCAATTAAATTGTTAACTACATTTTCTAG
AGAATCGTCAGGAACAGCTGATAAAGCGATCTTGAAATCTTCTTCTGACTGCATTGCAAGCCAAACTTTTTG
TGATAATTTAACATTTATGAACTCAGGACTCATAACTTGTTCAAAATATAAATCAAAGAATGCCGAATAAGC
AATCCTTCTATTTTTTAGGAATTCATTATTTGAATTTATATTATCAATATCAAATAAAACTTCTAGAAAAGA
CTCATACATTTCATTATCTTGAATAAAATCACTTAACTTAACTTTTCTTTTGTCATTATCTGATCGTGCCAA
GAGATAATCTTTAAGTTCAAAAATTTCTTTAAATTTATCTGGAAAGAAAATTCTTATCGCTTCAATAGTGAG
TAAATCAACCACATCAATTTCTTTACCTAATTGTTTAAAGATATTCGATAGAGAAGATGTGTAACGCTTAAT
ATCTCGAATATTTTTTATTGTTGGCTTAATGATATTCCAATATGCATTAGACCAACGCGCCTTATCTAGGTA
AACATCCCTTAAAATCTTATCTAAAGATGAAAATAAATTTTCTTGTAATAGTTTTTTAGGTACCTGTGGTAT
ATCGAATGGAATCTGAATTATCTTCTCTAAATAATCCTGGCCATCAATGGTATTATCATTTAATGGTTTAAT
TACTCTATTTTTATCAAATGATAAAACATAAACAATATTAGGAAAGTTTCCTGTAACTCTGACCAATTTTAG
AATTGATTGTAATTCATCAGATGATAAACGGTCTATATCATCTAAAATTACAGTAATAGGTTTACTTATTTC
CTTTAGAACTTTAATTAATTTATCACGTTGATTTTTCAAACTGTTTTTTTCTTTCTTTTTCTTTGAAAAAAA
ACTTAAACAGCCACCCAAGACACTAAAATAATTTCCTACAAATGGAATAGGTTTTAAATTAGATAACAACTC
TCCAAAACTACTCAAACTATCAATTAGCTCATTATCATCCTCATAATCTCTTAACTGAGCAGAGATTTCAGT
AAAAAATAAAGCAACTAAGTTATGAGCATCACTAAACATCCAAGGATTAAAATCAAGTACAAAAGAATTTTT
TTCTAATTCTGGTCGCATTAAATTTATATAGGATGTTTTACCATTTCCCCATTCTCCACATAATCCCACAAC
CAAACCTTCTTTATAGTCAAATGAAAAAATGTGTTTAGCAAATGCTTCTGCACTACTAGCTCTACCTAATAA
ATCATTGCTAGAATCTTTTATTGGATTATCGCTTATTAATTCCATATATTTTCCTTTAGTAAATGCTCATAT
CTTTTATGTGTAACC

SEQ ID NO:80 polynucleotide sequence comprising orfs26, 27 and non-coding flanking regions of these polynucleotide sequences.

TTATTGAATTTCCCTGGCAGAGAATAATATGACAAAAGTTTAGACAAAATTGCAAAACAATTAAGAGATTCT
GATAAAAAGGTTAATCTAATTTACGCCTTTAATGGAAGTGGAAAAACCCGTTTATCAAAAGTCTTTAAGAAT
CTTATTGCACCTAAAGAAAATCATGACAATGAAGAAGATCTAACACGAAGAAAAAATTCTTTATTTCAATGCC
TTTACCGAAGATTTATTCTATTGGGATAATGATCTACTTAATGACACAGAACCAAAATTAAAGATTCAACCA
AATTCTTTTATTCGCTGGTTGATTAGAGATCAAGGGGATGAAGGTAAAGTAATTGGAAAATTTCATCATTAT
TGTGATGAAAAACTTATGCCTAAATTTGATATAGAAAATAATCAAATTACATTCAGTTTTGCACGTGGAGAT
GATACGCCTGAAGAAAATATAAAACTATCGAAGGGGGAAGAAAGTAATTTTATTTGGAGTATTTTTCATACG
TTAATTGAACAAGTTGTTGCAGAATTAAATATCTCAGAGCCTAGTGAACGCACTACTAATGAATTTGATGAA
CTTAAATATATCTTTATTGATGATCCAGTAAGTTCATTGGATGAAAATCATCTTATTCAATTAGCTGTTGAT

TTAGCAGAATTAGTCAAAGATAGTCCCGATACTATAAAATTTATTATCACCACACACAATCCTTTATTTTAT
AACGTTTTATACAATGAACTTGGAGCAAAAAATGGTTATATTCTAAGAAAAGATGAAAATAAGAATGAAAAA
GAAAGATTTGATCTTGAGGTGAAACAAGGTGGTTCAAACAAGAGTTTCTCCTATCATCTTTTTCTAAAAAAT
CTACTTGAAGAAGTTGAACCTAAAGATATTCAAAAATATCACTTCATGTTACTGAGAAATTTATATGAAAAA
GCTGCTAACTTTCTTGGTTATTCAGGATGGTCAAATCTATTACCCAATGATGATGCAAGACAAAGCTATTAC
ACTCGTATAATCAATTTTACTAGTCACTCTACGTTATCAAATGAGATAATCGCTGAGCCAACAGATGCCGAA
AAGAAGATTGTTAAATATTTACTTGAACATCTAATTAATAATTATGGTTTCTATATAGAAGAAAATATTAAA
GACCCACAAACTGATAATATAACAGAGTAAAAATATGAACGACTTAATCATCTACAACACTGACGATGGTAA
ATCTCACGTTGCTTTATTAGTTATCGAAAATGAGGCTTGGCTGACTCAAAATCAGCTTGCGGAACTTTTTGA
CACCTCTGTACCAAATATAACCACTCATATAAAAAACATATTACAAGACAAAGAGTTAGATGAGTTTTCAGT
TATTAAGGATTACTTAATAACTGCCCAAGATAGCAAACAATATCAAGTAAAACATTATTCCCTTGATATGAT
TCTCGCCATCGGCTTTCGTGTGCGCAGCCCTCGTGGTGTACAGTTTCGTCGTTGGGCGAATACGCAATTACG
TACTTATTTAGATAAAGGTTTTCTATTAGATAAAGAGCGGTTGAAAAATCCTCAAGGTCGATTTGATCATTT
TGATGAATTACTGGAACAAATTCGCGAAATTCGAGCCAGTGAATTGCGGTTTTATCAAAAAGTACGAGAGTT
ATTTAAATTATCCAGTGACTACGATAAAACAGATAAAGTCACTCAAATGTTTTTTGCAGAAACACAAAATAA
GTTGATTTATGCCATTACACAACAAACCGCCGCAGAGCTTATTTGTACGCGTGCAAATGCCAAATTGCCTAA
TATGGGTCTTACCTCTTGGAAAGGTGCTGTTGTACGTAAAGGCGATATTATTACCGCTAAAAACTATTTAAC
TCATGATGAATTAGATTCTTTGAATCGTTTAGTGATGATCTTTTTAGAAAGTGCTGAATTACGCGTTAAAAA
TCGTCAAGATCTCACATTAAATTTCTGGCGTAATAATGTCGATAATTTAATTGAATTTAACGGTTTTCCGTT
GCTTATCGGTAATGGAACCCGAACCGTAAAACAAATGGAAACCTTTACCAAAGAACAATATGCCTTATTTGA
TCAGGTCAGAAAACAACAAAAACGCATACAAGCTGATAATGAAGATTTAGAAATTTTAGAAAACTGGCAGAA
AGATCTGAAAAAGCAAAAGCATTAAGGAACTACTT

SEQ ID NO:81 polynucleotide sequence comprising orfs28, 29 and non-coding flanking regions of these polynucleotide sequences.

AATTTTTCTACCCCCTCTTTCTCAAAGAGGGGGCAACCTGATAACATTATTTACATTCTAACCCGAGGACAT
CGTTTAAATTTTTCCCGTAAACTTATCATCATACCTAATCCACTGGAGATTGATGATGCCTTGGATAGAGAC
CGATGCGATGCAACAGCGTGTACTTTTTTTAAAAGCGTGGCTAAGCCAACGTTATACTAAAACTGAACTGTG
TCAGCAGTTTAATATTAGCCGTCCAACGGCAGATAAATGGATTAAACGCCACGAACAGCTTGGTTTTGAGGG
CTTAAGCGAGTTATCTCGTAAATCTTATCATAGCCCTAATGCCACGCCACAATGGATTTGTGACTGGCTTAT
CAGTGAGAAACTTAAACGTCCTCACTGGGGTGCCAAAAAGCTTTTAGATAACTTTACTCGGCATTTTCCAGA
AGCGAAAAAGCCGTCTGATAGCACGGGCGATTTAATTTTGGCGTGTGCAGGGTTAAAACGTCGTATGAGTGC
AGACACACAATCTTTTGGCGAATGCATCGCACCCAATACCACCTGGAGTGCTGACTTCAAGGGGCAATTTTT
ACTCGGCAATCAGAAGTTCTGCTATCCGCTGACGATTACAGATAATTTCAGTCGCTTTTTATTTTGTTGTAA
GGGGTTGCCGAATACAAAATCAGCGCCTGTTATTGCTGAGTTTGAACGTCTTTTTGAGCAATTTGGTCTGCC
GTATTCGATTCGTACCGATAACGATTCATCTTTTGCATCACAAGCATTAGGTGGATCTAGGTGTATTGACTT
AGGTATTCCTTCTGAACGAATTAAGCCATCACACCCAGAGCAGAACGGACGACACGAGCGAATGCACCGTAG
CTTAAAAACAGCGCTTCAACCTCAAAATAGCTTTGAAGCTCAACAGACATTCTTCAACCAATTCTTACGAGA
ATACAAAGAAGAATGTTCACACGAAGGCGTTTGACATATTTATTATCGCTTTTATTTACTGGGCAGTTTTGA
TGCTAAGGAAGTGAAAATTAAATCTGCCACACTGTGGCATAAATAATTTAATGAATGTAAACGATGTCCTTG
GGGGAGGTGCAAACTATGTTTGGGTTGTGTATCCCCTGCCGTGGCTAGTAATGTTCTGTCAACTCACTTCGA
CAGTGGTAATCTTGCTGAATTGTTTTCTTCTCATGCGCTACGGGTGAGCTCCGCTCTGATTTGACCGCTTAT
TTGTACCGCCAAAATTTCTTGGCTGCTCCTTAATGCATTTATTGCGCCGACTATATCATATTCTTTGTGATA
TATCTGCGACTTGGGTAATATCGGCTGGCATTTTTCGATGGGATAGTAAATGGATGTTTTTCATACTACGTA
ATTTGTAATCCAGTCACCGTCTGAACTCATGCCAAGATTGTGCTGAAGTTGAACGGTTTAAGTCTGATTTTT
GTTTCGCGTTTTACTGTATTTTCCGCATCTCCTTGGTAATTTGTTGCTTGCAAACTCTCAATATAAATCATT
GCGTGGTTTTTGCTGATTCGGTGTGGGATTTGATGCAAGTAGTTTTTTTTGTGGGTATTGGTGACTTTGTGG
TGCAATTTGGCGATTTTCGTATAACTGAACTTGACGCCATTATCTTGCTTATATTGTTCATTCTGCCAAGTT
AACCCGATTAAACATGAAGCGAGAATAGCCACAACGCTGCTTAATTCTGCGGATTTGTTCGCCGTTTGGCAT
TATTTCGAGCTTCAAGGCTCTGCGTAGTTGCATTGGCAAGGTTTAGGATATGATTTTCCTTATATTTTACTT
TTGGTCTATGAAAAAGAAATCCTCTTACTGTGGTGCATTCATTTTAATTATTTGCCAACACATCGAGCAACA
AAAACACCTGATTAGTTAGCTTTGAAACGGCTACGCCGTTGGTGTCTCATATCTCCGCCATGAAAGACGGAG
TTTTACGGCAGGAGGCT

SEQ ID NO:82 polynucleotide sequence comprising orfs30, 31, 32 and non-coding flanking regions of these polynucleotide sequences.

GGGTTGCCTGTTATAAACTATTAATTTTCTGATTGGTTATGTATATTTTTGCCATTTCTTCTAATTTGTTTA
CATCATCTTTATCATTAAAAATTGTTTTTTCATTTACAATTTTTGTCAAAATTAATTTCATTTCATTGTGGT
TTGAAGGTTTACAATAAGATAACACTAAATTTGCCCATTGAGTTATACGATCATCTTTAATGTAATTTCCCC
AAAGTTCAGTTAGAATATTTTCAGGAGTTTCTAAAATAAATTTTTTTAATTGCAAGAATATTGTTCTCTACT
CTCTTTAATACAGCAGAACAAAGATGTGTTTCACCACAAGTATCAGTTATTTTTTGTTGCCCTCCTGCAGAA

AACTCATCTCTTAAACTAGTGTTTATTACTCCATGTTTAGTCACTAGCCATAGTGCGAATTTATCATATTTA
TTTCTAGGATTTCCTAAGATCGTTTCAGGGAAGAAAGCATATGCTTGAGCAATTAACATATCTCGCTCATAT
TTTTCTATTACCTTCCAGTGTCTAACTTTGGGAGGGGCAATATCATCTAAAACATTGTTTGAAGTCCACCAT
AAACTTTCACCTTCTATTAATAGAGATTTATAGTATTTAGCTACAGGAGTTATTGGATTATCTAATTCTCGG
AGAGAATCATATGTGGTCTCCATTTTTTCAAATATGCTCTCCCCCTTTTCTAATAACATATCTATTTTATAT
CTAGGGTAATGGGTTACAGCTATATCACATAAACACGACTCATATGGTTTGGATAGGAATTTAATATTTCCA
CCTAATTTACCAAATGTTAAGAAAATTTTTTCTATATTTGGAATTCGTGTAGACTCAAGAATAGAACTGCCA
ATTGAAGTCCATTTATCTCCTTGTGTACTTTTTACTTCAATACCATAATATTGACTAGCTACAATGTCTGGA
AAATGTTTCCCTGATACTAAACTAATAGTGTCTTCGAAAGGAGTATTTTGAGCACAATAACAAATAGCCTCA
TATACATCTTTTTCTAAATCAATACCACTACGTTTCTTATAGTATGCAACCCTATTTTCTGCATCATGATTA
AGAAAATTATCGACTCTATTCATTAATGACGTGAATTCATGTAAAGGTGGATACTTATTTTTAGAGAAAATC
ATAAATAAATCCTATTTAAAATAAAGATTCCATTTTTTTTCTATATTTTAGCAATATTATAAGCTAATATA
CATGGTACCGCATTGCCAATAATTTTATAGGCATTACTGGCTGAAACAGAAACGTTTTCTGCTGTTTTAGGT
AAAATAAATTGGTATCTATCAGGAAACGTTTGTAATCTAGCACATTCTCTTATAGTAAGACGACGTTCGAGC
ATGCCTTTAGATAATTCGTTAATATATTTCCCTTCATGCTCTATGCTTAGCCTACGATTTTCAATATTACCA
TGATGTTCAGAATCGGAATTGTTGGGGCCCAACAGAAATTAAGTTTTAAATTTTCAAACCCTGGCCCCTTGG
ACCAATGGGTTTTCCCCCATAAATTATTTGGGGCTTTTGGGGAAATAATTTTTTGGTTTGAAAAAAGGGGGT
TCTTTTTGGTTATAAAAAATTGGGGGTTTCTTTTGGGAGGAATTTTATATTAAAAAGGGCCCTTTGGGGGCG
GCCATTGGGTAAACCCAACCCAGACTTTTC

SEQ ID NO:83 polynucleotide sequence comprising orf33 and non-coding flanking regions of these polynucleotide sequences.

ATGTTAAGGCTTGAGGCAAAGAATGGGCTCAAGCCTTTTGATTTCATCAAAATATAAAAATTAAGGAGATTA
TATGAGTGTACTCAGTTACGCACAAAAAATCGGTCAAGCCTTAATGGTGCCTGTGGCAGCCTTACCTGCTGC
TGCATTATTAATGGGTATTGGCTATTGGATCGACCCAGATGGTTGGGGTGCAAATAGTCAATTAGCCGCATT
ATTAATTAAATCTGGCGCAGCAATTATTGACAACATGGGCTTACTCTTCGCTGTGGGCGTCGCTTTTGGGCT
TGCAAAAGATAAACACGGTTCCGCCGCACTTTCAGGCCTTGTTGGTTTCTACGTAGTAACCACCCTACTTTC
CCCTGCTGGTGTAGCACAATTACAACACATTGATATTAGTGAAGTGCCTGCCGCATTCAAAAAAATCAATAA
CCAATTTATTGGGATTTTAATTGGTGTGATTTCAGCTGAACTTTACAACCGTTTCTATCAAGTTGAATTACC
AAAGGCACTTTCGTTCTTTAGCGGAAAACGCCTCGTCCCAATTTTGGTTTCTTTCGTGATGATCGCCGTATC
ATTTGCCTTACTCTATATTTGGCCTCATATTTTTAACGCTCTCGTTTCATTTGGTGAATCCATCAAAGATTT
AGGTGCAGTAGGTGCGGGGATCTACGGTTTCTTCAACCGCTTATTAATTCCTGTAGGCTTACACCATGCCTT
AAACTCTGTATTCTGGTTTGATGTAGCGGGTATCAACGATATTCCAAACTTCTTGGGCGGCGCTAAATCCAT
TGCCGAAGGCACTGCAACCGTGGGGCTAACTGGTATGTATCAAGCTGGTTTCTTCCCTGTCATGATGTTTGG
TTTACCAGGTGCTGCTCTTGCAATTTATCACTGCGCAAAACCAAACCAAAAAGTACAAGTGGCCTCAATTAT
GCTTGCGGGTGCGTTAGCCTCTTTCTTTACAGGGATCACTGAACCGCTTGAATTCTCATTTATGTTCGTTGC
ACCTGTACTTTATGTATTGCATGCATTATTAACAGGTATCTCTGTATTCATTGCAGCTACAATGCACTGGAT
TGCAGGATTCGGATTTAGTGCAGGTTTAGTGGATATGGTACTTTCTAGCCGTAACCCACTTGCCGTTAGCTG
GTATATGTTACTTGTACAAGGTATTGTATTCTTTGCTATCTATTATTTTGTGTTCCGTTTTGCAATTAATGC
CTTTAATCTCAAAACGCTAGGACGTGAAGATAAAGCGGAAACAGCTGCAGCCCCAACTCAAAGCGACCAATC
TCGCGAAGAAAGAGCGGTGAAATTTATTGCTGCTTTAGGTGGTTCAGAAAACTTCAAAACTGTGGATGCTTG
TATCACTCGTTTACGCTTAACTTTAGTTGATCATCACAATATTAACGAAGATCAACTTAAAGCGCTTGGTTC
AAAAGGTAATGTAAAATTAGGCAATGATGGATTACAAGTCATTTTAGGGCCTGAAGCTGAACTTGTGGCAGA
TGCG

SEQ ID NO:84 polynucleotide sequence comprising orf34 and non-coding flanking regions of these polynucleotide sequences.

GGGATTTCATTATGCTGTTTTACTTTATACTTTAAAAGTGCAAAAATAAAAAAACTCTTTTGCGCTAAACGG
AATAATAAAATGAAAACAACTTCTGAAGAATTAACGGTATTTGTGCAAGTAGTCGAAAATGGCAGTTTCAGC
CGTGCAGCCAAGCAGCTATCAATGGCAAATTCTGCGGTAAGTCGTGTGGTGAAAAGGCTAGAAGAAAAATTG
GGTGTGAACCTAATCAACCGCACTACTAGACAGCTTAGACTAACAGAAGAAGGCTTACAATATTTTCGTCGC
GTACAGAAAATTCTGCAAGATATGGCTGCAGCTGAAGCTGAAATGTTGGCAGTGCACGAAGTCCCACAAGGC
ATACTACGCGTAGATTCAGCCATGCCGATGGTGTTACATCTGCTAGTGCCACTGGCAGCAAAATTCAACGAA
CGCTATCCGCATATCCAACTTTCGTTAGTTTCTTCTGAAGGCTATATCAATCTGATAGAACGCAAAGTCGAT
ATTGCCTTACGAGCTGGAGAATTGGATGATTCTGGGCTGCGTGCTCGTCATCTATTTGATAGCCACTTCCGC
GTAATCGCCAGTCCAGACTACTTGGCAAAACACGGCACGCCACAATCAACTGAAGCTCTTGCCAACCATCAA
TGTTTAGGCTTCACTGAGCCCAGTTCACTAAATACATGGGAAGTTTTAGATGCTCAAGGAAATCCCTATAAA
ATCTCACCGTACTTTACCGCCAGCAGCGGTGAAATTTTACGGTCATTGTGTCTTTCAGGCTGTGGTATTGCT
TGCTTATCAGATTTTTTGGTAGACAATGACATCGCTGAAGGAAAATTAATTCCCTTACTTACTGAACAAACC
GCCAATAAAACGCTCCCCTTCAATGCTGTTTACTACAGCGATAAAGCAGTCAACCTTCGCCTACGTGTGTTT
TTAGACTTTTTAGTAGAAGAGCTAAGGGGATAATTAAAATTCATAGCATTGAATTTTAAAGTCAATTTGCAA

AAATACTTTAAAACCTGACCGCACTTGTCCCCCTGTCTTTTCATTACAATCTAGATTTCCTAACCTCCTTTC
AAAATCGCCCTCAATCTATCAAGTTGGTTTTGTGTTTTTTCTTGTTTTTGTT

SEQ ID NO:85 polynucleotide sequence comprising orf35 and non-coding flanking regions of these polynucleotide sequences.

CAGTTCATCATTGGGCTTTTTCATAAATTTATGAAAAAGGTAGAATAGCTGTTTTGTGGCGATAAAAAAAGA
CGCATTGAGCGTCTGTCTTTCCACCGCTCCAAGTTATTCAGAAACTGCGACATTCCCGACTTTCTGTTGAAA
GTGTGGTTATCTTAATCCGAAGTGAGGGCGGTGTCAAATAAAAAGCGCTGAGAATTTGAGGGAGCGAGTTAT
TCATCATCAATTAATTCTTTTGgTTTTCTTTGGAATGTCATTCACCTCTCCTTTAATACCATCAACAGCTTT
ATCCAGGCGTTTTCCTACTCCATCGATAATTGTTTCAAGTGGTGTGCTTTTTAAATCTTTGTCAAAGACTTT
GGTTGGATTTATCCCCAAATTATCCACGGCAATTTGCAGAAGTTGCTGATGTAATTTAGGGTCTTGTTCTTG
TACTTGTTTCTTATAACCTTCAAATGCCATTGCTGAGGAATATTTGTAGTTATAATCCTCCCTTAATCTAAA
GAGATAAGCCCGTTCTTTTGCTTTCAACCAGGCGATGACAAGTAACGGGATTGTCACAATGGACTTAGCAAG
AAATTGTAAAATATTAAGGCTGTCTGCTGCACTCAGGCTTGTTGAATAATTGAACAATGAAATAACAGATGT
TGCAACcAAGTGACCCcAAGgCAAAATTTTATCTACAGCTTTCATTTTACTATCGATATTTTCAGATTGAGT
TTTAAACGAACCTGCCATGCTTGCTCGGTTGGCGTCTTCAATAATCATTTCAATCTCCTCTTTTTGTTTATT
GAATAATTTAATCATACCTTCAATATCTTCATGATATTTTTcCGATTTGGGGTTTATTGGTTTTCCCGCTGT
GGTTGCTAATGTCGTAATTTTAGTAAGATTATTTTGTGCGGTGAATTCATAGTTCGAAATGTCGCCACTTAA
TTTCTCTGACTGTTCGTGCCACTGGGAAATTTCAGTTATTTGTTCTTGTGCGTCGTTATAAGATTTTTTGAG
TGTAATCAGTGAGTTTTTTAAATTTTCGAACTCCTTTTTATTCTCTACTAATGCTCTTCAAGTGAGATGTGG
TCTTCTAAATGGGGATCCTC

SEQ ID NO:86 polynucleotide sequence comprising orf36 and non-coding flanking regions of these polynucleotide sequences.

ATGAAAAGTTATTGCTATTATGCCTAAGCTAAAAACAAAATCCAGCATAAAAGCTGAATTTTTATGGATTGC
GTAGCATTATTGATTTAGTTGAAAACGATGCTTTTCAGGAATTAAAAATGACAAAAGCCACCTTTTAGGTGG
CCTTGTCTCAATATTGTAGGGGGGGGTGATAATGCTATCAGTGACCAACGTTCCCTATCGTCGGAGCGGAGT
CTATGGTAAAACAATTCAAATGTCAAGTGATAAGTAGGATTATATGTTATCAGCAACGCAATTTCTTGTTTT
AGAAAAAGCACTTAGTAAGGAAAGATTATCTACATACAAAAACTATGTGAAAAATAAAACTTCAGAAAGTAT
TAATGATAACATGGTTGCTTTATATGAATGGAATTCTGAAATAGCGGGCTATTTTCTTGAATTCTGTAATAT
ATATGAGATTTCATTAAGAAATGCTATTTATAGATCAATAGATTCGTATGATCATTATGGTATCAGACAGAG
ACAAATACTTAGACAAAGTCCTAAATTAAGAGAAAAAGTTGAAGAATTAGGTAGAAATGCGACTGATGGAAA
AATCATATCTAGTTTACATTTTCACTTTTGGGAATTTTTTGAAGAAGTTTTTCTTGTGGAATTCTCGTGAGC
TTCACAGAATGCCTCTTTTGTATGCTTATAGAATAATTTCTTTTGAAAACTCAAATAAAGATAAGGATATAT
TATTTATTATAAAAGTCACAAAGAATTTAAGAGTGAATATAAGAAACAGAATCTGTCATCACGATCCCATCT
TCAATAAAGATTTAAAGAAAATTCTGAAACAAGTTATGTGGGTATTTAGTAAAATTGATTATGATTTATACT
TAGTTATTAACAATCTATATTCCAATAAAATTATCAATCTTTTAAATAAGAAGCCAATCTGACTACAAATGT
AGAAGATCAGACCTCATCTGACAAATCACAATAAAAAATGAGCATTTCCTGTTTAGTATATGAGTGTCAAAC
TCAATCTAAACAGGAAATCCTCGTATTTTATTTTTACAACAGATTAG

SEQ ID NO:87 polynucleotide sequence comprising orf37 and non-coding flanking regions of these polynucleotide sequences.

GTATATCAATAGAGTATTTTTACAATATCATACTTTTAACTTATAATTCCAAACTAGATTATTATGGTCT
TAAACTGTTAGAAGAATATATATGATTGGAAAAAATCTTTATAACTATTGTTCTAACATTAACTCTAATT
AGGATATAAATGCACTTTTATCAATATCTAAACGCATTTCCATATGTAATTTCGGGGGATAAATGAAACT
AATATCTCTATTCTCAGGTTGTGGGGAATGGATATCGGATTTGAAGGTAATTTCTCTTGTCTAAAAAAA
TCTATTAATGAGGAGCTCCACCCTGAATGGATCAGCTCCACAGAAAATGAATGGGTTACCGTTTCGCCCA
CCTCTTTTGAGACAATTTTTGCTAATGATATTAAACCTGATGCTAAAGCAGCATGGGTTTCTTATTTCTT
AGACCAAAAAGCGAATGCAAACGAAATCTACCACTTAGAAAGCATTGTTGATCTTGTAAAAAAAGAACGG
GAAACTCACAATATTTTCCCAAAAGGCATTGATATATTAACAGGTGGATTTCCTTGTCAAGATTTTTCTG
TAGCCGGAAAACGATTAGGATTTGATTCTCACAAAAATCATCATGGAAAAATATCAAATATAGATGAACC
CTCAATTGAAAATAGAGGACAATTATACATGTGGATGAGAGAAGTAATATCTATAACTCACCCCAAATTA
TTCATAGCTGAAAATGTAAAAGGATTAACGAACCTTAAAGATGTAAAAGAAATTATTGAACATGATTTTG
GTCAAGCTAGTGACGAAGGATACTTAATTGTACCAGCTTCAGTATTAAATGCTCAGTTTTATGGAGCTCC
TCAATCACGTGAGCGTGTCATTTTTTTTTGGTTTTAAAAAAAAATGCGGCTAAAATAAAAAAAGCTTTTA
GAAGGAATTACCAAAAAGGAAAATATTGCCTGAGGAATTACCAATCCCTTATTCCTTCCCCCCAACTTCA
TGGGAAAAAGAAAAATTTTGAAAAGCCGGTTGGTACCTTGCCCCCCGATGGCTTTTAATAAATTCTCC